PARAMETRIZATION PROCEDURE SOLUTION OF OPTIMAL CONTROL PROBLEMS OVER SPLINE SPACES

A Thesis presented for the degree of
Doctor of Philosophy in Electrical Engineering
in the
University of Canterbury,
Christchurch, New Zealand.

by

F. S. Chou, B.Sc.(Hons.)

1978

ACKNOWLEDGEMENTS

For his patient guidance and invaluable advice throughout the period of this research, I am deeply indebted to my thesis supervisor, Dr. H. R. Sirisena.

I would like to thank the University Grants Committee of New Zealand for the award of a Postgraduate Scholarship.

Finally, I wish to thank Miss B. V. Nottingham for typing the thesis so expertly.

CONTENTS

ABSTRACT

In this thesis we report on investigations into the numerical solution of deterministic, continuous-time optimal control problems via parametrization techniques. These methods involve the linear expansion of one or more problem variables (i.e. control, state, co-state) in terms of basis functions. In this way the original optimal control problem can be reduced to a finite-dimensional minimization problem which may be solved numerically using standard mathematical programming algorithms.

Two specific techniques are considered here; we refer to them as the control parametrization (CP) and the state parametrization procedures (SP). As their names imply, the CP and SP procedures involve the expansion of the control and state, respectively, in terms of basis functions.

The importance of splines in the interpolation and approximation of functions is well known. In this research the viability of employing splines in conjunction with the above mentioned parametrization procedures is examined. The rates of convergence of the CP and SP solutions are also analysed; under appropriate smoothness conditions, explicit error bounds are derived for the control, state and cost functional convergences. Numerical results supporting the validity of these error bounds are presented.

All the numerical computations in this research have been done on the University of Canterbury Burroughs 6700/7700 machine using single-precision arithmetic.

PRINCIPAL NOTATIONS

ε	Is an element of
$\subset$	Is a subset of
$\cup$	Union
$\cap$	Intersection
inf	Infimum, or greatest lower bound
sup	Supremum, or least upper bound
E^n	n-dimensional Euclidean space
x^T	Transpose of x.
$\langle x,y \rangle$	Euclidean inner product of x and y.
$\|x\|$	Modulus of x
$\|\|x\|\|$	Norm of x
$\dot{x}$	Derivative of x with respect to time t.
$D^n x, x^{(n)}$	nth-derivative of x with respect to time t.
$C^\alpha[a,b]$	Space of functions with continuous αth order derivatives on $[a,b]$.
$PC^\alpha[a,b]$	Space of functions with piecewise continuous αth order derivatives.
S_h^α	Space of splines of order $\alpha-1$ with mesh size h.

CHAPTER 1

INTRODUCTION

1.1 NUMERICAL METHODS FOR OPTIMAL CONTROL PROBLEMS

The two major advances in modern control theory have been (a) the dynamic programming approach of Bellman [1] based upon the principle of optimality and (b) the maximum principle of Pontryagin [2] which is an extension of the classical calculus of variations.

The main result of the dynamic programming approach to the continuous time optimal control problem is the Hamilton-Jacobi-Bellman (HJB) equation for the optimal return function. In general this is a nonlinear partial differential equation which is extremely difficult to solve analytically. However, in those special cases when a closed form solution to the HJB equation can be found, the optimal control is obtained as a feedback law; that is, the optimal control is determined as a function of the state. The dynamic programming technique was originally developed for discrete time control problems. For these problems, application of the optimality principle results in a set of recursive relations which can be conveniently solved on a digital computer. The chief drawback of dynamic programming is its enormous computer storage requirements. This problem has been partly overcome by the state increment algorithm of Larson [3].

In comparison to dynamic programming, the maximum principle of Pontryagin provides a more popular approach to the numerical solution of the optimal control problem. Numerical methods based on the maximum principle can be generally classified as one of two types. The first category includes all the so-called indirect methods which are based on solving the two point boundary value problem (TPBVP) that results from the application

of the maximum principle to the optimal control problem. Some of the better known techniques belonging to this category are described below.

(1) Quasilinearization [4]: This method replaces the nonlinear TPBVP by a sequence of linear boundary value problems which can be solved easily. The major drawback of this method is that a good initial guess of the solution is usually necessary for convergence. However, if the method converges, it does so quadratically to the solution.

(2) Invariant imbedding [5]: This procedure imbeds the TPBVP within a class of more general initial value problems.

(3) Shooting method [6]: This method iterates on the initial values of the costate variables, leaving their terminal conditions unsatisfied until the solution is obtained. The main difficulty with this approach lies in the fact that boundary conditions are highly sensitive to small changes in initial conditions because the state-costate system of equations is inherently unstable.

The second category of numerical methods are the so-called direct methods which seek to prescribe a minimizing sequence of state and control functions without solving the TPBVP. For further details concerning these methods the text by Sage [7] can be consulted. Some important direct methods are described below.

(1) Gradient method: Also known as the method of steepest descent, this is a control iterative procedure in which corrections to the control are made in the direction of most rapid change in the cost functional. The algorithm is based on calculating the first order effects of the control on the cost functional, and is therefore a first order method. Initial convergence is usually excellent, but in a vicinity of the optimum convergence becomes slow. Two variants of the basic gradient method are the min-H and conjugate gradient methods.

The min-H method was proposed by Kelley [8] and proceeds as follows: guess a nominal control, then solve for the state and costate. Keeping the state and costate fixed, determine the control which minimizes the Hamiltonian; this control is then used for the next iteration. Further developments of this method can be found in [51].

The conjugate gradient method was originally developed by Lasdon et. al. [9] as an extension of the conjugate gradient method of Fletcher and Reeves [10] in finite-dimensional space. This algorithm is reported to possess superior convergence properties to the basic gradient method.

(2) The method of second variations: This is an extension of the gradient method based on considering the second-order effects of the control on the cost functional as well as the first-order ones, resulting in greatly improved convergence characteristics near the optimum. However, in addition to being a more complicated algorithm than the gradient method, it has a smaller region of convergence and requires a reasonably good (convex) nominal solution [48].

(3) The mathematical programming approach [11]: This procedure replaces the original optimal control problem by a static optimization problem, and constitutes the subject of our present investigation. We now review the method in greater detail.

1.2 THE MATHEMATICAL PROGRAMMING APPROACH

The development of mathematical programming (MP) is in an advanced stage: a well established theory is in existence and a wide range of highly sophisticated computational techniques are available for the numerical solution of MP problems. To exploit the theoretical and computational sophistication of MP for the purpose of solving optimal control problems

efficiently constitutes the primary objective of the MP approach. The procedure consists of two steps. Firstly, an optimal control problem is reformulated as an MP problem via a discretization scheme. The resulting MP problem is then solved by means of a suitable computational algorithm.

Among the earliest to apply this approach to the solution of optimal control problems are Zadeh and Whalen [12] in 1962. Early applications of the MP approach generally employ some finite difference scheme to discretize the original control problem, but many papers have since appeared proposing various discretization schemes. Some of these schemes will be discussed later. But for the moment we shall review some MP techniques.

Review of Mathematical Programming

The components of an MP problem are:

(a) a scalar objective function $\phi(x)$ of the vector variable x,

(b) a vector function $g(x)$ representing inequality constraints, and

(c) a vector function $h(x)$ representing equality constraints.

The statement of the MP problem is:

$$\text{minimize } \phi(x), \text{ subject to } g(x) \leq 0,\ h(x) = 0. \tag{1.1}$$

An MP problem is called a linear programming problem when its objective function ϕ and constraints g and h are all linear; otherwise it is called a nonlinear programming problem. An important special case of the latter is the quadratic programming problem in which g and h are linear while the objective function ϕ is quadratic. Other special classes of MP problems include geometric programming, integer programming and convex programming (see [13]).

Numerous algorithms are available for the numerical solution of MP problems. Some of these are applicable to the general MP problem while others are applicable only to special classes of MP problems. Among the

numerical techniques of MP, the more important ones include,

(1) the Simplex algorithm [14] for linear programming,

(2) Wolfe's algorithm [15] for quadratic programming,

(3) the method of feasible directions [16],

(4) the gradient projection method [17], [18] of Rosen, and

(5) the sequential unconstrained minimization technique, originally proposed by Carroll [19] and developed further by Fiacco and McCormick [20].

Some of the above methods involve intermediate solutions of unconstrained minimization problems. Numerical techniques for unconstrained minimization are divided into two groups: direct search methods and descent methods.

(1) Direct search methods require the evaluation of function values but not their gradients. Well known algorithms belonging to this group include,

(a) pattern search method [21] of Hooke and Jeeves,

(b) Rosenbrock's method of rotating coordinates [22], and

(c) Powell's method [23].

(2) Descent methods require the evaluation of gradients as well as function values. In general, descent methods are more efficient than direct search methods. At each iteration of a descent technique a descent direction is computed and the minimum of the objective function is then searched in that direction. Each descent method is characterized by the particular manner in which these search directions are generated. Examples of descent methods are,

(a) method of steepest descent,

(b) second order gradient method,

(c) method of conjugate gradients, and

(d) variable metric method.

Of the examples listed above, the steepest descent method is the simplest. The search directions used here are the locally steepest directions. Computations with this method have proved to be rather unsatisfactory. This poor performance can be attributed to the fact that the steepest descent method converges asymptotically in a two-dimensional space (see [24]).

The overall performance of the second order gradient method is quite good, and its convergence is particularly good in the vicinity of the minimum. However, the computational load for each iteration is quite large.

The method of conjugate gradients was first used by Hestenes and Steifel [25] to solve systems of linear equations by minimizing the corresponding quadratic objective function. This was later generalised to general functions by Fletcher and Reeves [10]. This method is efficient and does not suffer from the drawback mentioned above for the steepest descent method.

The variable metric method was proposed by Davidon and extended by Fletcher and Powell [27]. The method exhibits convergence properties similar to that of the second order gradient method.

Discretization Schemes

The initial step in the implementation of the MP approach to an optimal control problem is to select a discretization scheme. Using this scheme, a finite-dimensional optimization problem is obtained as an approximation to the original optimal control problem. In general, different discretization schemes will lead to different MP problems and therefore to different approximate solutions of the optimal control problem. For a discretization scheme to be viable, the solution of the MP problem resulting

from the application of the scheme to the original problem must be reasonably close to the true solution. Moreover, the approximate solution should improve as the discretization is gradually refined. Other features of a discretization scheme to be considered are the ease of implementation of the scheme and the complexity of the resulting MP problem. Examples of discretization schemes for optimal control problems include,

(1) finite difference,

(2) control parametrization,

(3) Ritz-Trefftz,

(4) trajectory approximation,

(5) Ritz-Galerkin.

In Chapter 4 of this thesis, an alternative discretization scheme based on parametrizing some components of the state vector will be introduced. We shall refer to this procedure as,

(6) state parametrization.

Of the above mentioned examples, the finite difference scheme (see Tabak and Kuo [11]) is the simplest. It involves the discretization of the time variable : the entire time interval under consideration is divided into sub-intervals of equal size. Throughout each sub-interval of time, the control and state are assumed to be constant. The state equation is replaced by a finite difference equation, and the cost functional is approximated by a finite summation which is to be minimized with respect to the piecewise constant control.

Each of the remaining procedures (2) to (6) involves the parametrization of one or more of the following quantities : (a) control, (b) state and (c) co-state.

For instance, the control parametrization procedure [28], [49] involves making the following approximation on the control variable:

$$u(t) = \sum_{i=1}^{m} q_i \psi_i(t) , \qquad (1.2)$$

where $\psi_1,\ldots,\psi_m$ are known basis functions of time and $q_1,\ldots,q_m$ are unknown coefficients. The control is said to be parametrized by $q_1,\ldots,q_m$ because once these are specified, the control function can be found from (1.2) The corresponding state vector can be determined by integrating the state equation, which in turn means that the cost functional J can be evaluated. Thus, the original optimal control problem reduces to that of finding the optimal values of $q_1\ldots,q_m$ which minimize the cost J.

The Ritz-Trefftz procedure [29] involves the parametrization of the co-state, while the trajectory approximation procedure [30] involves the parametrization of both the state and co-state. In the Ritz-Galerkin procedure [31] the control, state and co-state are all parametrized simultaneously. A more detailed summary of these procedures can be found in Appendix B.

A procedure closely related to the Ritz-Galerkin procedure is described by Neuman and Sen [32], whereby the control and state are simultaneously parametrized and required to satisfy a weakened version of the state equation through the use of the collocation procedure.

In their paper [26] on the generalized gradient method for optimal control problems with state inequality constraints and singular arcs, Mehra and Davis pointed out that instead of treating the control variable as the independent variable all the time, it is sometimes advantageous to treat a state variable as the independent variable. This point of view is emphasised in the state parametrization procedure to be introduced in Chapter 4.

We now consider the choice of basis functions for the parametrization procedures. Generally, standard families of functions like the polynomials, trigonometric functions and Legendre polynomials can be employed. Computational results using polynomials are reported in [32], while Chebyshev polynomials are employed in [33].

A class of functions that have received considerable attention in recent times are the spline functions. These are piecewise polynomials with continuity requirements only slightly less stringent than the polynomials. Nevertheless, spline functions possess certain desirable properties not shared by polynomials which make them particularly attractive for use in interpolation and approximation (see [50]). In this thesis we shall be primarily concerned with the application of spline functions in the control parametrization and state parametrization procedures.

In a series of articles [29], [31], [34] Bosarge et. al. examined the convergence properties of the Ritz-Trefftz and Ritz-Galerkin procedures. In particular, explicit error bounds which indicate the rates at which the approximate solutions converge to the true solutions are derived in the case where spline functions are employed. Extensions of this work to the control parametrization and state parametrization procedures are contained in Chapters 3 and 5 of this thesis.

1.3 DISTRIBUTED PARAMETER SYSTEMS

Up to now, our discussion has been concerned with control problems involving systems described by ordinary differential equations, known as lumped parameter systems, or simply, lumped systems. However, there are many problems occurring in physical applications which involve systems described by partial differential equations. Such systems are known as

distributed parameter systems, or simply, distributed systems, in which the state variables are dependent on one or more spatial coordinates in addition to time.

Among the first papers to appear in this area have been those of Butkovskii and Lerner [35] and Butkovskii [36]. These early efforts have been directed towards generalising the theory for lumped problems to accommodate distributed problems. Specifically, a maximum principle for distributed problems was developed in the above mentioned articles. Pioneering work in this field has also been done by Wang [37], who extended the fundamental concepts of controllability, observability and stability to the distributed case. Subsequently, much work has been done towards putting the control theory of distributed systems onto a solid mathematical foundation. In particular, Balakrishnan [38] and Lions [39] have formulated and analysed distributed control problems in abstract settings (viz. Banach and Hilbert spaces) using the tools of functional analysis. In spite of the significant progress achieved in recent years, the theoretical development of distributed control is far from complete, especially in regard to nonlinear problems.

From a computational point of view, the solution of distributed control problems represents a much more formidable task than the solution of lumped control problems with increased programming complexity as well as increased computer time and memory requirements. Hence the development of efficient numerical algorithms for solving distributed problems is of great importance. We now review some of the available numerical techniques for distributed problems.

One popular approach to developing numerical methods is to extend existing methods for solving lumped problems. Thus, the method of descent,

the conjugate gradient method and the method of second variations have all been extended to various classes of distributed problems, (see [40], [41]).

Another popular approach is to reformulate the distributed problem as a lumped problem. This can be achieved using a finite differencing technique [42]. For problems involving spatially independent control variables, the Galerkin procedure can also be used to obtain the approximate lumped problem (see [43], [44]).

Bosarge et. al. [45], [46] have presented the Ritz-Trefftz and Ritz-Galerkin procedures for distributed problems involving parabolic systems. In Chapter 6 we develop the state parametrization procedure for distributed problems as an extension of the procedure introduced in Chapter 4 for lumped problems.

1.4 THESIS ORGANIZATION

The chapter headings for the remainder of the thesis together with an abstract for each chapter are as follows:

Chapter 2: The Control Parametrization Procedure.

The CP procedure for solving optimal control problems is reviewed, with special emphasis on the use of spline functions. The transformation technique for converting control problems involving control or state variable inequality constraints into problems without constraints is described. Computational results employing spline approximation spaces are reported.

Chapter 3: Error Bounds for the Control Parametrization Procedure.

Convergence of the CP procedure is examined. Error estimates for the approximate control, state and cost functional over arbitrary finite-dimensional approximation spaces are obtained in the L_2-norm. For the CP approximations over spline spaces, explicit order bounds are derived. Computational results supporting these error bounds are presented.

Chapter 4: The State Parametrization Procedure.

The SP procedure for solving optimal control problems is developed. The procedure is then specialised to the class of control problems whose state equations can be expressed in the phase variable form. Computational results employing cubic splines are presented for two specific examples.

Chapter 5: Error Bounds for the State Parametrization Procedure.

Convergence of the SP procedure is examined. Error bounds for the SP approximations over spline approximation spaces are established for a class of control problems.

Chapter 6: The State Parametrization Procedure for Distributed Parameter Systems.

The SP procedure is extended to a wide class of distributed control problems. Computational results using multivariate splines in conjunction with the SP procedure are presented. The convergence of the procedure is also examined.

Chapter 7: General Conclusions.

The contributions of this thesis are summarised and directions for further research are suggested.

REFERENCES

[1] R. Bellman: Dynamic Programming, Princeton University Press, Princeton, N.J., (1957).

[2] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze and E. F. Mishchenko: The Mathematical Theory of Optimal Processes, Wiley, N.Y., (1962).

[3] R. E. Larson: State Increment Dynamic Programming, American Elsevier, N.Y., (1968).

[4] R. Bellman and R. Kalaba: Quasilinearization and Nonlinear Boundary Value Problems, Elsevier Press, N.Y., (1965).

[5] E. S. Lee: Quasilinearization and Invariant Imbedding, Academic Press, N.Y., (1968).

[6] P. B. Bailey and L. F. Shampine: On Shooting Methods for TPBVP, J. Math. Anal. and Appl., Vol.18 (1967) pp.45-58.

[7] A. P. Sage: Optimum Systems Control, Prentice-Hall, Englewood Cliffs, N.J., (1968).

[8] H. J. Kelley: Method of Gradients, in Optimization Techniques, G. Leitman ed., Academic Press, (1962).

[9] L. S. Lasdon, S. K. Mitter and A. D. Waren: The Conjugate Gradient Method for Optimal Control Problems, IEEE Trans. Automatic Control, AC-12 (1967) pp.132-138.

[10] R. Fletcher and C. Reeves: Function Minimization by Conjugate Gradients, Computer J., Vol.7 (1964) pp.149-154.

[11] D. Tabak and B. C. Kuo: Optimal Control by Mathematical Programming, Prentice-Hall, Englewood Cliffs, N.J., (1971).

[12] L. A. Zadeh and B. H. Whalen: On Optimal Control and Linear Programming, IRE AC-7 (1962) pp.45-46.

[13] S. Vajda: Mathematical Programming, Addison-Wesley, Reading, Mass. (1961).

[14] G. B. Dantzig: Linear Programming and Extensions, Princeton University Press, Princeton, N.J., (1963).

[15] P. Wolfe: The Simplex Method for Quadratic Programming, Econometrica, Vol.27 (1959) pp.382-398.

[16] G. Zoutendijk: Methods of Feasible Directions, Elsevier, Amsterdam, (1960).

[17] J. B. Rosen: The Gradient Projection Method for Nonlinear Programming I: Linear Constraints, SIAM J. Appl. Math., Vol.8 (1960), pp.181-217.

[18] J. B. Rosen: The Gradient Projection Method for Nonlinear Programming II: Nonlinear Constraints, SIAM J. Appl. Math., Vol.9 (1961) pp.514-532.

[19] C. W. Carroll: The Created Response Surface Technique for Optimizing Nonlinear Restrained Systems, Op. Res., 9 (1961).

[20] A. Fiacco and G. McCormick: Nonlinear Programming: Sequential Unconstrained Minimization Techniques, Wiley, N.Y. (1968).

[21] R. Hooke and T. A. Jeeves: Direct Search Solution of Numerical and Statistical Problems, Journal of the ACM, Vol.8 (1961), pp.212-229.

[22] H. H. Rosenbrock: An Automatic Method for Finding the Greatest or Least Value of a Function, Computer J., Vol.3 (1960), pp.175-184.

[23] M. J. D. Powell: An Efficient Method for Finding the Minimum of a Function of Several Variables without Calculating Derivatives, Computer J., Vol.7 (1964) pp.147-151.

[24] J. Kowalik and M. R. Osborne: Methods for Unconstrained Optimization Problems, Elsevier, N.Y., (1968).

[25] M. Hestenes and E. Stiefel: Method of Conjugate Gradients for Solving Linear Systems, Report 1659, National Bureau of Standards, (1952).

[26] R. K. Mehra and R. E. Davis: A Generalized Gradient Method for Optimal Control Problems with Inequality Constraints and Singular Arcs, IEEE Trans. Automatic Control, AC-17 (1972) pp.69-79.

[27] R. Fletcher and M. J. D. Powell: A Rapidly Convergent Descent Method for Minimization, Computer J., Vol.6, (1963).

[28] H. H. Rosenbrock and C. Storey: Computational Techniques for Chemical Engineers, Pergamon, London (1966).

[29] W. E. Bosarge, Jr. and O. G. Johnson: Direct Method Approximation to the State Regulator Problem Using a Ritz-Trefftz Suboptimal Control, IEEE Trans. Automatic Control, AC-15 (1970) pp.627-631.

[30] L. L. Lynn, E. S. Parkin and R. L. Zahradnik: Near-Optimal Control by Trajectory Approximation, Ind. Engng. Chem. Fund., Vol.9 (1970) pp.58-63.

[31] W. E. Bosarge, Jr., O. G. Johnson, R. S. McKnight and W. P. Timlake: The Ritz-Galerkin Procedure for Nonlinear Control Problems, SIAM J. Numer. Anal., Vol.10 (1973), pp.94-111.

[32] C. P. Neuman and A. Sen: Weighted Residual Methods in Optimal Control, IEEE Trans. Automatic Control, AC-19 (1974) pp.67-69.

[33] G. J. Lastman: Suboptimal Open Loop Control of Nonlinear Systems Using Approximations for the Controls, Int. J. Control, Vol.20 (1974), pp.289-303.

[34] W. E. Bosarge, Jr., and O. G. Johnson: Error Bounds of High Order Accuracy for the State Regulator Problem via Piecewise Polynomial Approximations, SIAM J. Control, Vol.9 (1971), pp.15-28.

[35] A. G. Butkovskii and A. Ya. Lerner: The Optimal Control of Systems with Distributed Parameters, Automation and Remote Control, Vol.21 (1960), pp.472-477.

[36] A. G. Butkovskii: The Maximum Principle for Optimum Systems with Distributed Parameters, Automation and Remote Control, Vol.22 (1962), pp.1156-1169.

[37] P. K. C. Wang: Control of Distributed Parameter Systems, Advances in Control Systems, Academic Press, N.Y.(1964).

[38] A. V. Balakrishnan: Optimal Control Problems in Banach Spaces, SIAM J. Control, Vol.3 (1965).

[39] J. L. Lions: Optimal Control of Systems Governed by Partial Differential Equations, Springer-Verlag, Berlin (1971).

[40] J. H. Holliday and C. Storey: Numerical Solution of Certain Nonlinear Distributed Parameter Optimal Control Problems, Int. J. Control, Vol.18 (1973), pp.817-825.

[41] D. E. Cornick and A. N. Michel: Numerical Optimization of Distributed Parameter Systems by the Conjugate Gradient Method, IEEE Trans. Automatic Control, AC-17 (1972) pp.358-362.

[42] A. P. Sage and S. P. Chaudhuri: Gradient and Quasilinearization Computational Techniques for Distributed Parameter Systems, Int. J. Control, Vol.6 (1967) pp.81-98.

[43] C. P. Neuman and A. Sen: A Rapid Sub-Optimal Control Algorithm for Distributed Parameter Regulator Problem, Int. J. Control, Vol.16 (1972) pp.539-548.

[44] L. L. Lynn and R. L. Zahradnik: The Use of Orthogonal Polynomials in the Near-Optimal Control of Distributed Systems by Trajectory Approximation, Int. J. Control, Vol.12 (1970), pp.1079-1087.

[45] W. E. Bosarge, Jr., O. G. Johnson and C. L. Smith: A Direct Method Approximation to the Linear Parabolic Regulator Problem over Multivariate Spline Bases, SIAM J. Numer. Anal., Vol.10 (1973) pp.35-49.

[46] R. S. McKnight and W. E. Bosarge, Jr.: The Ritz-Galerkin Procedure for Parabolic Control Problems, SIAM J. Control, Vol.11 (1973) pp.510-524.

[47] G. Strang and G. J. Fix: An Analysis of the Finite Element Method, Prentice-Hall, N.J. (1973).

[48] A. E. Bryson Jr. and Y. C. Ho: Applied Optimal Control, Blaisdell, Waltham, Mass. (1969).

[49] T. J. Walder and C. Storey: Numerical Solution of an Optimal Temperature Problem, The Chemical Engineering J., Vol.1 (1970) pp.120-128.

[50] J. H. Ahlberg, E. N. Nilson and J. L. Walsh: The Theory of Splines and Their Applications, Academic Press, N.Y. (1967).

[51] R. G. Gottlieb: Rapid Convergence to Optimum Solutions using a Min-H Strategy, AIAA J., Vol.5 (1967) pp.322-329.

CHAPTER 2

THE CONTROL PARAMETRIZATION PROCEDURE

2.1 INTRODUCTION

The primary purpose of the present chapter is to review the control parametrization (CP) procedure for solving optimal control problems with special emphasis on the use of spline functions.

Suppose that we have an optimal control problem requiring the minimization of a cost functional $J(u)$ over the space of admissible controls U. The CP procedure discretizes the problem by restricting the control variable to a subset C_m of U which is parametrized by m real variables. This involves specifying the members of C_m to be of the form,

$$u(t) = F(q,t) \tag{2.1}$$

where F is a known function of the m-dimensional parameter vector $q \varepsilon E^m$. The function F specifies the form of the parametrization.

If only controls belonging to C_m are considered, then the cost functional reduces to a function of q, and we write,

$$J(u) = \tilde{J}(q) \tag{2.2}$$

The original goal of minimizing $J(u)$ over U is now replaced by that of minimizing $\tilde{J}(q)$ over C_m. Assuming the existence of a solution, the new problem of determining the optimal values of $q_1,\ldots,q_m$ is generally a much simpler task than the original problem which involves optimization in function space.

The special case when F is a linear function of q is of particular importance; equation (2.1) then takes the form

$$u(t) = \sum_{i=1}^{m} q_i \psi_i(t) \tag{2.3}$$

where $\psi_1, \ldots, \psi_m$ are known functions of time called basis functions. In this case C_m becomes an m-dimensional linear space over the scalar field E, and we call C_m an approximation subspace of U.

Sometimes the basis functions can be specially constructed to suit the problem in hand (see [1], [2]). Otherwise standard families of functions may be used, some examples being the polynomials, trigonometric functions, Legendre polynomials, Bessel functions and Hermite polynomials. Extensive computational results using Chebyshev polynomials have been reported by Lastman [16].

Hicks and Ray [2] presented computational results employing a control parametrization consisting of a bang-bang portion followed by a polynomial curve. The parameters to be optimized in this case are (a) the switching times for the bang-bang portion and (b) the coefficients for the polynomial curve. The static optimization was performed using the direct search technique of Rosenbrock. However, this proved to be rather unsatisfactory, as different nominal controls tended to converge to different final controls, even though the final values of the cost were in close agreement with one another. This difficulty appears to be due to the flatness of the cost surface in the vicinity of the optimum, which causes direct search methods to be severely affected by truncation and round-off errors in computing the cost.

In view of this problem, Sirisena [3] suggested that the static optimization could be more efficiently performed by a descent method. He

also proposed a spline parametrization of the control; this work was later extended to problems involving constraints [4].

In section 2.3 we review the basic CP procedure for the unconstrained optimal control problem. Extension of the procedure to problems with terminal constraints is reviewed in section 2.4. In sections 2.5 and 2.6 the transformation technique for converting problems involving control or state variable inequality constraints into unconstrained problems is reviewed. Throughout, the use of spline functions has been emphasised. The fundamentals of spline functions will now be reviewed.

2.2 REVIEW OF SPLINE FUNCTIONS

A spline function (or spline) is a piecewise polynomial which satisfies a strong smoothness condition. It is a natural generalisation of the polynomial, yet it possesses certain features not shared by polynomials which make it attractive for the purposes of interpolation and approximation. For a comprehensive treatment of spline functions refer to the text by Ahlberg, Nilson and Walsh [5]. In this section we review some basic facts about spline functions. We shall begin with a couple of definitions.

Definition: Consider an interval $[a,b]$ on the real line and a strictly increasing sequence of real numbers $t_0,t_1,\ldots,t_N$, where

$$a = t_0 < t_1 < \ldots < t_N = b.$$

The set $P = \{t_0,t_1,\ldots,t_N\}$ is said to be a partition of $[a,b]$, and the elements of P are called knots (or nodes). The mesh size h of P is defined by

$$h = \max \{t_{j+1} - t_j \mid j = 0, 1,\ldots,N - 1\}. \qquad (2.4)$$

Definition: Given a partition $P = \{t_o, t_1, \ldots, t_N\}$ of the interval $[a,b]$, a spline function $s(t)$ of degree α with knots $t_o, t_1, \ldots, t_N$ is a function defined on $[a,b]$ having the following properties:

(i) in each sub-interval (t_i, t_{i+1}), where $i = 0,1,\ldots,N-1$, $s(t)$ is a polynomial of degree α or less.

(ii) $s(t)$ and its derivatives of order up to $\alpha-1$ are continuous in (a,b).

It is clear from the above definition that a spline function of degree α is a function in $C^{\alpha-1}[a,b]$ whose derivative of order α is piecewise constant.

Given a partition $P = \{t_o, t_1, \ldots, t_N\}$, let us now consider the set of all splines of degree α over the partition P. It is easy to see that such a set is a linear space over the field of real numbers. Now, each spline $s(t)$ in the space is continuously differentiable $(\alpha-1)$ times in $[a,b]$, and the α^{th} derivative of $s(t)$ is a constant in each sub-interval (t_i, t_{i+1}), $i=0,1,\ldots,N-1$. Hence $s(t)$ is parametrized by the initial values of $s(t)$ and its first $(\alpha-1)$ derivatives together with the value of its α^{th} derivative over each sub-interval (t_i, t_{i+1}), $i=0,1,\ldots,N-1$. Therefore, the dimension of this linear space is $N+\alpha$, and henceforth it shall be denoted by $S^{\alpha+1}(t_o,\ldots,t_N)$, $S_N^{\alpha+1}$ or $S_h^{\alpha+1}$.

Spline Bases

It follows from the theory of linear spaces that there exists a basis containing $N+\alpha$ elements for the spline space $S^{\alpha+1}(t_o,\ldots,t_N)$. Moreover, this basis is not unique. To construct one such basis, we note that every $s(t) \in S^{\alpha+1}(t_o,\ldots,t_N)$ can be expressed in the form

$$s(t) = p_\alpha(t) + \sum_{i=1}^{N-1} q_i (t-t_i)_+^\alpha \qquad (2.5)$$

where $p_\alpha(t)$ is a polynomial of degree α or less, and $(t-t_i)_+^\alpha$ is the truncated power function defined as follows:

$$(t-t_i)_+^\alpha = \begin{cases} (t-t_i)^\alpha & \text{if } t \geq t_i \\ 0 & \text{if } t \leq t_i \end{cases} \tag{2.6}$$

Therefore the set $\{1,t,\ldots,t^\alpha,(t-t_1)_+^\alpha,\ldots,(t-t_{N-1})_+^\alpha\}$ is a basis for $S^{\alpha+1}(t_o,\ldots,t_N)$.

The B-splines of Schoenberg [6] provide another possible spline basis. In spline interpolation problems, the use of B-splines gives rise to better numerical stability than the use of bases constructed with truncated power functions (see [9]). A feature of the B-spline is that it has finite support; that is, it is non-zero only on a finite region. In fact, it was shown by Curry and Schoenberg [7] that the B-splines are of minimal support; that is, no other basis functions can be found which have smaller support regions than those of the B-splines. Another property of the B-spline is that it is strictly positive within its support region which is restricted to $\alpha+1$ adjacent intervals.

Let us consider now the bi-infinite partition $\{t_i\}_{i=-\infty}^{\infty}$ of the real line. The B-spline $\psi_i(t)$ of degree α which is non-zero over the interval $(t_i,t_{i+\alpha+1})$ is given by the formula,

$$\psi_i(t) = (t_{i+\alpha+1}-t_i) \sum_{j=1}^{i+\alpha+1} \{(t_j-t)_+^\alpha / \prod_{\substack{m=i \\ m\neq j}}^{i+\alpha+1} (t_j-t_m)\}. \tag{2.7}$$

The above definition provides the normalized form of the B-spline; in other words,

$$\sum_{i=-\infty}^{\infty} \psi_i(t) = 1 \tag{2.8}$$

for each fixed t.

In the case of a uniform partition, the B-splines of order α are merely translates of one another along the real line. To illustrate this point, suppose that the knots in the partition are given by $t_j = jh$ for all integers j, where h is the constant mesh size of the partition. Then it can be shown that,

$$\psi_i(t) = \psi_j(t-t_{i-j}). \tag{2.9}$$

Furthermore, a change in the mesh size of the uniform partition simply means a re-scaling of the independent variable in the formula for B-splines. Consider the partition $Z \equiv \{i\}_{i=-\infty}^{\infty}$ whose nodes are the integers, and let $\psi(t)$ be the B-spline of order α which is non-zero over the interval $[i, i+\alpha+1]$. Suppose $\{t_i\}_{i=-\infty}^{\infty}$ is another uniform partition, where $t_i = ih$, and let $\phi(t)$ be the B-spline of order α which is non-zero over the interval $[jh, (j+\alpha+1)h]$. The functions ψ and ϕ are then related by the formula

$$\phi(t) = \psi(\frac{t}{h} - j+i) \tag{2.10}$$

The B-spline $\psi(t)$ starts from zero at $t = i$, monotonically increases till it reaches its maximum at the point $t = i + (\alpha+1)/2$, then monotonically decreases till it returns to zero at $t = i + \alpha + 1$. Moreover, ψ is symmetrical about its point of maximum. Explicit formulas for typical B-splines of orders 1, 2 and 3 over the partition Z are given below.

(i) linear spline

$$\psi(t) = \begin{cases} 1 + t, & t \in [-1,0] \\ 1 - t, & t \in [0,1] \\ 0 & \text{elsewhere} \end{cases}$$

(ii) parabolic spline

$$\psi(t) = \begin{cases} \frac{1}{2}(1+t)^2, & t \in [-1,0] \\ \frac{1}{2} + t - t^2, & t \in [0,1] \\ 2 - 2t + \frac{1}{2}t^2, & t \in [1,2] \\ 0, & \text{elsewhere} \end{cases}$$

(iii) cubic spline

$$\psi(t) = \begin{cases} \frac{1}{6}(2+t)^3, & t \in [-2,-1] \\ \frac{2}{3} - t^2 - \frac{1}{2}t^3, & t \in [-1,0] \\ \frac{2}{3} - t^2 + \frac{1}{2}t^3, & t \in [0,1] \\ \frac{1}{6}(2-t)^3, & t \in [1,2] \\ 0 & \text{elsewhere} \end{cases}$$

2.3 PROBLEMS WITHOUT CONSTRAINTS

In this section we review the CP procedure for the unconstrained Bolza problem. Although spline approximation spaces are employed here, the method to be described can be generalised to the case of a general parametrization with obvious modifications. For clarity of presentation, only single-input systems will be considered; the generalisation to multivariable systems is straightforward.

Problem Statement

Consider the dynamical system described by

$$\dot{x}(t) = f(x(t), u(t), t), \quad x(0) = x_o \tag{2.11}$$

and the cost functional

$$J(u) = \theta[x(T),T] + \int_0^T \phi(x(t), u(t), t)\, dt, \tag{2.12}$$

where x(t) is an n-dimensional state vector and u(t) is a scalar-valued control variable belonging to the set of admissible controls U. The optimal control problem is to find the admissible control u* that minimizes the cost functional:

$$J(u^*) = \inf\{J(u) \,|\, u \in U\}. \tag{2.13}$$

Method of Approach

Firstly, a suitable partition of the time interval $[0,T]$ is chosen. For convenience, we shall assume that the partition is uniform:

$$0 = t_o < t_1 < \ldots < t_{N-1} < t_N = T$$

where N is the number of sections into which $[0,T]$ is divided by the partition, and $t_j = jh$ $(0 \leq j \leq N)$, where $h = T/N$ is the uniform mesh size.

An arbitrary spline function u^h of order α over the partition (2.13) may be expressed in the form

$$u^h(t) = \sum_{i=1}^{N+\alpha} q_i \psi_i(t) \ , \tag{2.14}$$

where $\{\psi_1, \psi_2, \ldots, \psi_{N+\alpha}\}$ is a basis for the spline space $S_h^{\alpha+1}$. In general, U will be the space of piecewise continuous controls, so $S_h^{\alpha+1} \subset U$ for non-negative integer α.

The CP procedure involves finding the control $\bar{u}^h \in S_h^{\alpha+1}$ which minimizes the cost functional:

$$J(\bar{u}^h) = \inf\{J(u^h) \,|\, u^h \in S_h^{\alpha+1}\}. \tag{2.15}$$

It is clear that since $S_h^{\alpha+1}$ is a subset of U, we must have that $J(u^*) \leq J(\bar{u}^h)$, so in general $\bar{u}^h$ will only be sub-optimal.

For a given basis $\{\psi_1, \psi_2, \ldots, \psi_{N+\alpha}\}$, the control u is parametrized by the $N + \alpha$ parameters $q_1, q_2, \ldots, q_{N+\alpha}$. Hence the functional J(u) may be considered an ordinary function of these parameters, and we write

$$J(u) \equiv \tilde{J}(q_1, q_2, \ldots, q_{N+\alpha}). \tag{2.16}$$

It was recommended by Sirisena [3] that the optimization of $\tilde{J}$ be performed using a gradient search method, for instance the Davidon-Fletcher-Powell algorithm [8]. The implementation of this algorithm

requires the evaluation of the gradient of $\tilde{J}$ with respect to each parameter q_i. These gradients may be computed as described below.

The Hamilton H is defined in the usual manner,

$$H(x(t),u(t),\lambda(t),t) \equiv \phi(x(t),u(t),t) + \langle\lambda(t),f(x(t),u(t),t)\rangle. \tag{2.17}$$

The co-state equations are given by

$$\dot{\lambda} = -\frac{\partial H}{\partial x}, \quad \lambda(T) = \frac{\partial\theta[x(T),T]}{\partial x(T)}, \tag{2.18}$$

and by standard variational arguments, it can be shown that

$$\frac{\partial\tilde{J}}{\partial q_i} = \int_0^T \frac{\partial H}{\partial u}\psi_i(t)dt, \quad i = 1,2,\ldots,N+\alpha \tag{2.19}$$

Remarks

(i) As the number of sections N is increased, we do not necessarily have a monotonic decrease in the minimum cost $J(\bar{u}^h)$, unless the partition is successively refined.

(ii) The internal knots $t_1,t_2,\ldots,t_{N-1}$ of the partition may be assumed to be variable, and treated as additional parameters to be optimized. Expressions for the gradients of $\tilde{J}$ with respect to these knots have been derived in [3]. It is clear that whereas the control is a linear function of the parameters $q_1,q_2,\ldots,q_{N+\alpha}$, it is a nonlinear function of the variable knots $t_1,t_2,\ldots,t_{N+\alpha}$. If the optimal control profile is smooth, there is probably not much to be gained from using variable knots. However, if there are discontinuities in the optimal control profile, it would almost certainly be profitable to optimize the internal knot locations as well.

(iii) The parametrization introduced by Sirisena in [3] implicitly employs a basis which is closely related to the truncated power functions; the basis functions are of the form

$$\psi_i(t) = t^{i-1}/(i-1)!, \qquad i = 1,2,\ldots,\alpha$$

$$\psi_{i+\alpha}(t) = \left[(t-t_{i-1})_+^{\alpha} - (t-t_i)_+^{\alpha}\right]/\alpha!, \qquad i = 1,2,\ldots,N$$

2.4 PROBLEMS WITH TERMINAL CONSTRAINTS

In this section we briefly review the CP procedure for problems in which the state vector is constrained at the final time. The procedure to be described here is due to Sirisena and Tan [4].

The problem statement of section 2.3 is assumed, with the addition of the following terminal constraint:

$$\eta[x(T),T] = 0, \tag{2.20}$$

where η is an ℓ-dimensional vector.

Method of Approach

As in the procedure for the unconstrained problem, the first step is to select a suitable partition. The control is then parametrized as a spline function as in equation (2.14).

The terminal constraint (2.20) is treated as ℓ equality constraints on the parameters $q_1,\ldots,q_{N+\alpha}$. The optimization subject to equality constraints can be handled using numerical techniques like the gradient projection algorithm of Rosen. Further details concerning the implementation of the CP procedure may be found in [4].

Remarks

Given a particular problem, and assuming that N and α are fixed, must there always exist a control in $S_h^{\alpha+1}$ such that the terminal constraint $\eta[x(T),T] = 0$ is satisfied? Intuitively, we feel that, provided N and α are sufficiently large, such a control should indeed exist.

We can show this to be true for a constant linear system. Suppose that the following system

$$\dot{x} = Ax + bu \tag{2.21}$$

is controllable, where A is an $n \times n$ constant matrix and b is a constant

n-dimensional vector. (The dependence of x and u on t have not been explicitly specified above for the sake of notational convenience).

Then the augmented system

$$\begin{aligned} \dot{x} &= Ax + bu_1 \\ \dot{u}_1 &= u_2 \\ &\vdots \\ \dot{u}_\alpha &= u_{\alpha+1} \end{aligned} \tag{2.21a}$$

with $u_{\alpha+1}$ as the new control variable can be shown to be controllable by applying the usual controllability criterion, (see [11]).

Suppose now that we require $u_{\alpha+1}$ to be piecewise constant. This is equivalent to the introduction of sampling to the system (2.21a). The effect of sampling on the controllability of a continuous linear system has been investigated by Kalman, Ho and Narendra [10] for the case of constant sampling period. It was found that, provided the sampling frequency is sufficiently large, there should be no loss of controllability. Hence the system (2.21) is controllable using only elements of $S_h^{\alpha+1}$ provided that the mesh size h is sufficiently small.

Notation

Henceforth we shall simplify our notations by omitting the argument list after each function whenever it is convenient to do so without causing confusion.

2.5 PROBLEMS WITH CONTROL CONSTRAINTS

Optimal control problems occurring in practice often involve inequality constraints on the state and control variables. These constraints take on the general form,

$$C(x,u,t) \lesssim 0. \tag{2.22}$$

When the function C is independent of x, the constraint is then called a control constraint:

$$C(u,t) \lesssim 0.$$

By far the most important class of control constraints is that of saturation-type constraints of the form

$$c(t) \lesssim u(t) \lesssim d(t) \tag{2.23}$$

In this section we demonstrate that the CP procedure is applicable to problems involving saturation-type constraints through the use of suitable transformations of variables. Problems involving state constraints are considered in section 2.6.

Problem Statement

We consider here the problem of minimizing the cost functional

$$J = \int_0^T \phi(x,u,t)dt \tag{2.24}$$

for the n-dimensional system

$$\dot{x} = f(x,u,t), \qquad x(o) = x_o \tag{2.25}$$

subject to the control constraint (2.23). For the sake of clarity we shall assume C and u to be scalar-valued functions.

Method of Approach

The transformation technique we use here involves defining the control variable u in terms of a new variable ν by means of a suitable function which ensures that the control constraint is automatically satisfied for all possible values of the new variable ν. The original problem with control constraint may then be replaced by an unconstrained problem in the new control variable ν. We can then apply the CP procedure to this new unconstrained problem.

Thus, for the constraint (2.23) we define the new variable ν by

$$u(t) = \theta(\nu,t) \ , \tag{2.26}$$

where θ is any function that satisfies the condition

$$c(t) \leq \theta(\nu,t) \leq d(t) \tag{2.27}$$

for all possible values of ν.

The CP procedure then proceeds with the parametrization of the new control variable

$$\nu(t) = F(q,t) \tag{2.28}$$

where q is the m-dimensional parameter vector.

We now need to derive the gradient expression of the cost $\tilde{J}(q)$ with respect to q. Using standard variational arguments, we can show that

$$\frac{\partial \tilde{J}}{\partial q} = \int_0^T \frac{\partial H}{\partial \nu} \frac{\partial \nu}{\partial q} \, dt \ , \tag{2.29}$$

where H is the Hamiltonian defined in the usual way,

$$H \equiv \phi + \langle \lambda, f \rangle \ , \tag{2.30}$$

and λ is the co-state vector given by

$$\dot{\lambda} = - \frac{\partial H}{\partial x} \ , \quad \lambda(T) = 0 \tag{2.31}$$

Finally, using (2.26) and (2.28), the expression (2.29) may be written

$$\frac{\partial \tilde{J}}{\partial q} = \int_0^T \frac{\partial H}{\partial u} \frac{\partial \theta}{\partial \nu} \frac{\partial F}{\partial q} \, dt. \tag{2.32}$$

Types of Transformation

We now look at two possible forms of the transformation θ. The first example is one which is well known in the context of mathematical programming (see [12]):

$$\theta(\nu,t) = \frac{d(t) + c(t)}{2} + \frac{d(t) - c(t)}{2} \sin \nu(t). \tag{2.33}$$

However, the computation of transcendental functions is quite expensive, and a computationally simpler alternative is the following transformation:

$$\theta(\nu,t) = \begin{cases} d(t), & \text{if} \quad \nu(t) > d(t) \\ \nu(t), & \text{if} \quad c(t) \leq \nu(t) \leq d(t) \\ c(t), & \text{if} \quad \nu(t) < c(t) \end{cases} \tag{2.34}$$

The main draw-back of the above "clipping" transformation is that convergence to the optimal solution is not always obtained. For instance, if the nominal control lies completely outside the feasible region, the CP procedure employing the transformation (2.34) breaks down. Nevertheless, in those instances when it does work, it appears to work well. Highly satisfactory computational results using the "clipping" transformation in conjunction with the CP procedure have been reported by Sirisena and Tan [4].

Numerical Example

The CP procedure described above was used to solve the following Rayleigh problem which was taken from [13].

For the system described by

$$\dot{x}_1 = x_2 \qquad , \quad x_1(0) = -5$$

$$\dot{x}_2 = -x_1 + 1.4x_2 - 0.14x_2^3 + 4u \quad , \quad x_2(0) = -5$$

find the control u* that minimizes

$$J = \int_0^{2.5} (x_1^2 + u^2)\,dt$$

subject to the control magnitude constraint

$$|u(t)| \leq 1 \quad \text{for all } t \in [0,2.5].$$

The transformation (2.33) was employed to reduce the above problem to an equivalent unconstrained problem in the new control variable ν given by

$$u = \sin \nu \ .$$

The variable ν was parametrized as a parabolic spline over a uniform partition of $[0,2.5]$ with N sections. The problem was solved for N=2 and N=4; and for each N used, two nominal controls were employed, viz. $\nu \equiv 0$ and $\nu \equiv 1.55$. It was found that the control converged to the same value after starting from the two different nominal controls. The converged control profiles for N=2 and N=4 are summarised in Table 2.1. The minimum cost obtained for N=2 was $J(\bar{u}) = 42.805$, and for N=4, the minimum cost $J(\bar{u}) = 42.790$.

TABLE 2.1 CONTROL PROFILES

t	N=2 $\bar{u}$	N=4 $\bar{u}$	t	N=2 $\bar{u}$	N=4 $\bar{u}$
0.0	.9966E0	.1000E1	1.3	.5698E0	.5909E0
0.1	.9998E0	.1000E1	1.4	.4182E0	.3975E0
0.2	.9998E0	.1000E1	1.5	.2766E0	.2244E0
0.3	.9993E0	.1000E1	1.6	.1521E0	.8438E-1
0.4	.9994E0	.1000E1	1.7	.4879E-1	-.1659E-1
0.5	.1000E1	.1000E1	1.8	-.3149E-1	-.7686E-1
0.6	.9993E0	.1000E1	1.9	-.8806E-1	-.9739E-1
0.7	.9945E0	.1000E1	2.0	-.1209E0	-.1004E0
0.8	.9810E0	.9985E0	2.1	-.1303E0	-.9489E-1
0.9	.9533E0	.9889E0	2.2	-.1162E0	-.8089E-1
1.0	.9047E0	.9597E0	2.3	-.7858E-1	-.5834E-1
1.1	.8281E0	.8944E0	2.4	-.1722E-1	-.2724E-1
1.2	.7166E0	.7744E0	2.5	.6778E-1	.1244E-1

2.6 PROBLEMS WITH STATE CONSTRAINTS

We considered the use of appropriate transformation in section 2.5 to handle control constraints. This approach is extended here to problems involving state constraints of the form (2.22) in which the constraint function C contains the state x explicitly.

The system equation (2.25) and the cost functional (2.24) are assumed here. We begin by defining a new variable z through the transformation

$$C(x,u,t) = \theta(z,t) \ , \tag{2.35}$$

where $\theta(z,t)$ is a negative function. We assume here that C is linear in x and u. This assumption ensures that the form of the constraint function C does not implicitly impose constraints on z. And it was for the same reason that we only considered saturation-type control constraints in the previous section. Fortunately, the linearity of C is not a restrictive condition, as most constraints appearing in control problems are formulated as linear constraints.

Having specified the transformation (2.35), we can then obtain the transformed (unconstrained) problem following the procedure described in [14] by Jacobson and Lele. In that article, the authors suggested using the particular transformation

$$\theta(z,t) = -\frac{1}{2}z^2 \tag{2.36}$$

Nevertheless, any negative function θ may be used, and the procedure described in [14] regarding the conversion of the original problem to the new unconstrained problem applies to the general case with obvious modifications.

Basically, there are two cases to consider. If C contains u explicitly, equation (2.35) can be solved for u in terms of x and z, and u can be eliminated from the original optimal control problem. The result is an unconstrained problem in the new control variable z. If C does not contain u explicitly, then the equation (2.35) is repeatedly differentiated with respect to t until u appears. Suppose that the constraint (2.35) is of order p; that is, u appears for the first time in d^pC/dt^p. Then the original problem converts into an unconstrained problem of increased dimension n+p, with the new control variable d^pz/dt^p.

Numerical Examples

The method described above was applied to two sample problems. The transformation of Jacobson and Lele were employed to reduce each problem to an unconstrained one. And in each case the new control variable was parametrized as a linear spline.

Example 1.

Minimize $J = \frac{1}{2}\int_0^1 u^2 \, dt$

for the system

$$\dot{x} = u \, , \quad x(o) = 0$$

subject to the first-order constraint

$$x(t) \geq \sin(\pi t) + a \, ,$$

where $a \equiv \pi\sqrt{3}/12 - .5 = -.046550$ (to 5 significant figures).

This problem has been taken from [15], and has the exact solution

$$u^*(t) = \begin{cases} 6a + 3, & 0 \leq t \leq 1/6 \\ \pi\cos(\pi t), & 1/6 \leq t \leq 1/2 \\ 0 \, , & 1/2 \leq t \leq 1 \end{cases}$$

The minimum cost for the problem is

$$J^* = 1.0992 \text{ (to 5 significant figures)}$$

Employing the transformation (2.36), we have

$$x(t) - \sin(\pi t) - a = \frac{1}{2} z^2(t).$$

By differentiating the above equation and making use of the state equation, we obtain

$$u = zz_1 + \pi\cos(\pi t) \ ,$$

where $z_1 \equiv \dot{z}$. From the above defining equation for z, we find that $z(0) = \sqrt{-2a}$ (we could have chosen the initial condition $z(0) = -\sqrt{-2a}$, but it does not matter which one is used).

The transformed problem is then:

$$\text{minimize } J = \frac{1}{2} \int_0^1 \left[zz_1 + \pi\cos(\pi t)\right]^2 dt$$

for the system

$$\dot{x} = zz_1 + \pi\cos(\pi t) \ , \quad x(o) = 0$$

$$\dot{z} = z_1 \ , \quad z(0) = \sqrt{-2a} \ .$$

Because the state x is not present in the performance index, the corresponding state equation is actually now redundant.

Our new control variable is z_1 , and we parametrized it as a linear spline over a uniform partition of $[0,1]$ with N sections. A solution was obtained for N=5, and the minimum cost obtained was

$$J(\bar{z}_1) = 1.1014.$$

The approximate solution obtained is summarised in Table 2.2, while the exact solution is summarised in Table 2.3.

TABLE 2.2 SOLUTION PROFILES (N=5)

t	$\bar{u}(t)$	$\bar{x}(t)$	$\bar{z}_1(t)$
0.0	.2575E1	.0000E0	-.1857E1
0.1	.2796E1	.2733E0	-.1302E1
0.2	.2508E1	.5422E0	-.7480E0
0.3	.1850E1	.7625E0	-.3239E0
0.4	.9688E1	.9047E0	.1003E0
0.5	.1078E0	.9556E0	.1627E1
0.6	-.8481E-2	.9511E0	.3153E1
0.7	-.3021E-1	.9494E0	.2971E1
0.8	-.3349E-1	.9457E0	.2789E1
0.9	-.7726E-1	.9398E0	.2501E1
1.0	-.4458E-1	.9328E0	.2213E1

TABLE 2.3 SOLUTION PROFILES (OPTIMAL)

t	$u^*(t)$	$x^*(t)$
0.0	.2721E1	.0000E0
0.1	.2721E1	.2721E0
0.2	.2542E1	.5412E0
0.3	.1847E1	.7625E0
0.4	.9708E0	.9045E0
0.5	.0000E0	.9535E0
.	.	.
.	.	.
.	.	.
1.0	.0000E0	.9535E0

Example 2

$$\text{Minimize } J = \int_0^1 (x_1^2 + x_2^2 + .005\ u^2)dt$$

for the second-order system

$$\dot{x}_1 = x_2 \qquad x_1(0) = 0$$

$$\dot{x}_2 = -x_2 + u, \qquad x_2(0) = -1$$

subject to the first-order constraint

$$x_2(t) - 8(t-.5)^2 + .5 \leq 0.$$

This problem has been taken from [14]. Application of the transformation (2.36) gives rise to the following unconstrained problem

$$\dot{x}_1 = x_2 \qquad x_1(0) = 0$$

$$\dot{x}_2 = -zz_1 + 16t - 8 \qquad x_2(0) = -1$$

$$\dot{z} = z_1 \qquad z(0) = -\sqrt{5}$$

$$J = \int_0^1 x_1^2 + x_2^2 + .005(x_2+16t-8-zz_1)^2\ dt.$$

The new control variable z_1 was parametrized as a linear spline over a uniform partition of $[0,1]$ containing N sections. The problem was solved for N=5, and the results are summarised in Table 2.4. The minimum cost obtained was

$$J(\bar{z}_1) = 0.17177.$$

The above problem was solved by Lastman [16] using Chebyshev polynomials in conjunction with the CP procedure. He obtained the minimum cost 0.17401 using 5 parameters, and 0.17038 using 10 parameters.

TABLE 2.4 SOLUTION PROFILES (N=5)

t	$\bar{u}(t)$	$\bar{x}(t)$	$\bar{z}_1(t)$
0.0	.1237E2	-.1000E1	.9555E1
0.1	.4441E1	-.1347E0	.8115E1
0.2	-.6757E0	.3208E-1	.6675E1
0.3	-.3002E1	-.1877E0	.3098E1
0.4	-.2017E1	-.4200E0	-.4792E0
0.5	-.5027E0	-.5003E0	-.1060E0
0.6	.1184E1	-.4201E0	.2671E0
0.7	.2459E1	-.1936E0	.3323E1
0.8	.6634E0	.8821E-2	.6379E1
0.9	-.2369E0	.1424E-1	.5374E1
1.0	.4750E0	.1257E-1	.4370E1

2.7 CONCLUSIONS

In this chapter we have reviewed the CP procedure for solving optimal control problems with fixed final time. The basic procedure was first described for the unconstrained problem. It was then seen that the procedure could be applied to problems involving terminal constraints because such constraints can be treated as equality constraints on the parameters. It was also shown that problems with control or state variable inequality constraints could be handled through the use of appropriate transformations of variables to convert these constrained problems into unconstrained ones.

We have also reviewed some basic material on spline functions, and the use of splines in the CP procedure has been emphasised. Finally, some results of computations employing splines in conjunction with the CP procedure have been presented.

REFERENCES

[1] T. J. Walder and C. Storey: Numerical Solution of an Optimal Temperature Problem, The Chemical Engineering J., Vol.1 (1970) pp.120-128.

[2] G. A. Hicks and W. H. Ray: Approximation Methods for Optimal Control Synthesis, Can. J. Chem. Eng., Vol.49 (1971) pp.522-528

[3] H. R. Sirisena: Computation of Optimal Controls Using a Piecewise Polynomial Parametrization, IEEE Trans. Automatic Control, AC-18 (1973) pp.409-411.

[4] H. R. Sirisena and K. S. Tan: Computation of Constrained Optimal Controls Using Parametrization Techniques, IEEE Trans. Automatic Control, AC-19 (1974) pp.431-433.

[5] J. H. Ahlberg, E. N. Nilson and J. L. Walsh: The Theory of Splines and Their Applications, Academic Press, N.Y. (1967).

[6] I. J. Schoenberg: Contributions to the Problem of Approximation of Equidistant Data by Analytic Functions, Quart. Appl. Math., Vol.4 (1946), Pt A, pp.45-99; Pt B, pp.112-141.

[7] H. B. Curry and I. J. Schoenberg: On Polya Frequency Functions. IV., J. Analyse Math., Vol.17 (1966) pp.71-107.

[8] R. Fletcher and M. J. D. Powell: A Rapidly Convergent Descent Method for Minimization, Computer J., Vol.6 (1963) pp.163-168.

[9] T. N. E. Greville: Introduction to Spline Functions, in Theory and Applications of Spline Functions, ed. T.N.E. Greville, Academic Press, N.Y. (1969).

[10] R. E. Kalman, Y. C. Ho and K. S. Narendra: Controllability of Linear Dynamical Systems, in Contributions to Differential Equations, Vol.1, Interscience Publishers, John Wiley & Sons, Inc. (1964).

[11] A. P. Sage: Optimum Systems Control, Prentice-Hall, Englewood Cliffs, N.J. (1968).

[12] M. J. Box: A Comparison of Several Current Optimization Problem Methods and the Use of Transformation in Constrained Problems, Computer J., Vol.9 (1966) pp.67-76.

[13] D. H. Jacobson: New Second-Order and First-Order Algorithms for Determining Optimal Control: A Differential Dynamic Programming Approach, J. Optimiz. Theory Appl., Vol.2 (1968) pp.411-440.

[14] D. H. Jacobson and M. M. Lele: A Transformation Technique for Optimal Control Problems with a State Variable Inequality Constraint, IEEE Trans. Automatic Control, AC-14 (1969) pp.457-464.

[15] W. W. Hager and G. Strang: Free Boundaries and Finite Elements in One Dimension, Mathematics of Computation, Vol.29 (1975) pp.1020-1031.

[16] G. J. Lastman: Sub-optimal Open Loop Control of Nonlinear Systems Using Approximations for the Controls, Int. J. Control, Vol.20 (1974) pp.289-303.

CHAPTER 3

ERROR BOUNDS FOR THE CONTROL PARAMETRIZATION PROCEDURE

3.1 INTRODUCTION

In the CP procedure described in the previous chapter, we seek the control function that minimizes the cost functional over a known subset of the space of admissible controls. The control determined in this way will only be sub-optimal in general, unless the optimal control happens to lie in the subset concerned. However, it is clear that more accurate approximations can be generated by expanding this subset. It is the purpose of this chapter to determine the rate at which the CP approximation approaches the optimal solution as the subset is expanded. For technical reasons we shall only be concerned with linear approximations, in which the form of the control parametrization is linear in the undetermined parameters.

Error estimates for the Ritz-Trefftz and Ritz-Galerkin procedures have been derived by Bosarge et.al. [1], [2]. In this chapter we derive similar error estimates for the CP procedure. This is done for unconstrained problems in sections 3.2 and 3.3, and for problems with fixed terminal state in section 3.4. By specializing to spline approximation spaces, relevant order bounds in terms of the mesh parameter h are also obtained.

3.2 UNCONSTRAINED PROBLEMS - LINEAR QUADRATIC CASE

In this section we carry out an error analysis of the CP procedure for unconstrained linear quadratic problems. The error analysis will be extended to general nonlinear problems in the next section, but the linear

quadratic case is considered first because the derivation of error bounds for this case is simpler and much neater.

Problem Statement

Consider the optimal control problem described by the time-varying linear system,

$$\dot{x}(t) = A(t)\, x(t) + B(t)\, u(t) , \quad x(0) = x_o , \tag{3.1}$$

and the quadratic cost functional

$$J(u) = \int_0^T \left[\langle x(t), Q(t)\, x(t) \rangle + \langle u(t), R(t)\, u(t) \rangle \right] dt. \tag{3.2}$$

We assume that x(t) is an n-dimensional state vector, u(t) is an r-dimensional control vector, A is an n × n matrix, B is an n × r matrix, Q is an n × n symmetric, positive semi-definite matrix, and R is an r × r symmetric, positive definite matrix. It is also assumed that A, B, Q and R are $PC^{\alpha-1}[0,T]$, where $\alpha \geq 1$. That is, the $(\alpha-1)$th derivative of each element of the above matrices is piecewise-continuous in the time interval $[0,T]$. The final time T is assumed to be fixed.

Under the above assumptions on the smoothness of the problem, it can be shown by standard differential equation arguments (see [4]) that $u^* \in \{PC^{\alpha}[0,T], E^r\}$ and $x^* \in \{PC^{\alpha}[0,T], E^n\}$. Hence we can take the space of admissible controls U to be the Sobolev space $\{W_2^{\alpha}[0,T], E^r\}$, and the state space X to be $\{W_2^{\alpha}[0,T], E^n\}$ (see Appendix A for the definition of Sobolev spaces).

Remarks

Let the functional $||.||_R$ be defined by

$$||u||_R \equiv \left[\int_0^T \langle u, Ru \rangle dt \right]^{\frac{1}{2}} .$$

Using the assumption that R is positive definite, it may be verified that $||.||_R$ is a norm on U and that there exist positive constants γ_1 and γ_2 such that

$$\gamma_1||u||_2^2 \leq ||u||_R^2 \leq \gamma_2||u||_2^2 \tag{3.3}$$

for every $u\varepsilon U$. Similarly, the functional $||.||_Q$ is defined by

$$||x||_Q \equiv \left[\int_0^T \langle x, Qx\rangle dt\right]^{\frac{1}{2}} .$$

However, as Q is only assumed to be positive semi-definite, $||.||_Q$ is not strictly a norm. Nevertheless, it can be shown that there exists a positive constant γ_3 such that

$$||x||_Q^2 \leq \gamma_3||x||_2^2 \tag{3.4}$$

for every $x\varepsilon X$.

Some Properties of Linear Systems

Let (u,x) be a control-state pair satisfying the state equation and initial condition (3.1). Then

$$\dot{x}(t) - \dot{x}^*(t) = A(t)\left[x(t) - x^*(t)\right] + B(t)\left[u(t) - u^*(t)\right],$$

$$x(0) - x^*(0) = 0. \tag{3.5}$$

By applying the Gronwall Inequality [6] to (3.5), it can be shown (see [3]) that

$$||x - x^*||_2^2 \leq K\,||u - u^*||_2^2 , \tag{3.6}$$

where K is a positive constant that will in general depend on the system matrices A and B.

It is also demonstrated in [3] that the quadratic cost functional J(u) defined by (3.2) satisfies the following identity

$$J(u) = J(u^*) + ||x - x^*||_Q^2 + ||u - u^*||_R^2 . \tag{3.7}$$

The inequality (3.6) and identity (3.7) will be used in the sequel.

Finite-Dimensional Approximation Spaces

Let $C_m \subset U$ be the finite-dimensional linear space spanned by the functions $\psi_1, \psi_2, \ldots, \psi_m$ where $\beta\psi_i \in U$ for each $i = 1,2,\ldots,m$ and each $\beta \in E^r$. Thus

$$C_m \equiv \{u \mid u = \sum_{i=1}^{m} \beta_i \psi_i , \quad \beta_i \in E^r\}. \tag{3.8}$$

In the CP method, the original problem of minimizing $J(u)$ over the function space U is replaced by the minimization of $J(u)$ over the finite dimensional space C_m. We seek the control $\bar{u}_m \in C_m$ such that

$$J(\bar{u}_m) = \inf \{J(u_m) \mid u_m \in C_m\}. \tag{3.9}$$

A word regarding notation: we say that C_{m+1} is the linear span of the basis $\{\psi_1, \psi_2, \ldots, \psi_{m+1}\}$, but $\psi_1, \psi_2, \ldots, \psi_m$ need not coincide with the basis functions that generate C_m. In other words, it is not assumed that $\{C_m\}$ is an expanding sequence with $C_m \subset C_{m+1}$. However, we do assume that the set $C \equiv \lim_{m\to\infty} C_m$ is dense in U, in the sense that for each admissible control u and for each $\delta > 0$, there exists $\nu \in C$ such that $||u - \nu||_2 < \delta$.

Now it is well known that the normed linear space $H_2 = \{L_2[0,T], E^r\}$ is a uniformly convex Banach space (see [5]). Since C_m is a finite-dimensional subspace of H_2 and $u^* \in H_2$, we know from approximation theory (see Appendix A) that there exists a unique best approximation $\hat{u}_m$ from C_m to u^*; i.e.,

$$||\hat{u}_m - u^*|| = \inf\{||u_m - u^*|| \mid u_m \in C_m\}. \tag{3.10}$$

Note, however, that in general $\hat{u}_m$ is not the same as the CP approximation $\hat{u}_m$ defined by (3.9).

Error Estimates

It is clear that

$$J(\bar{u}_m) \leq J(\hat{u}_m) \ , \tag{3.11}$$

and making use of identity (3.7), we have that

$$J(\bar{u}_m) - J(u^*) \leq ||\hat{x}_m - x^*||_Q^2 + ||\hat{u}_m - u^*||_R^2 \tag{3.12}$$

Applying (3.3), (3.4) and (3.6) to (3.12), we obtain the following error estimate for the cost $J(\bar{u}_m)$

$$J(\bar{u}_m) - J(u^*) \leq (\gamma_2 + \gamma_3 K) \ ||\hat{u}_m - u^*||_2^2 \ . \tag{3.13}$$

Combining the identity (3.7) with (3.13), it is clear that

$$||\bar{u}_m - u^*||_R^2 \leq (\gamma_2 + \gamma_3 K) \ ||\hat{u}_m - u^*||_2^2 \ , \tag{3.14}$$

and using the left-hand part of inequality (3.3), we have

$$||\bar{u}_m - u^*||_2^2 \leq \gamma_1^{-1}(\gamma_2 + \gamma_3 K) \ ||\hat{u}_m - u^*||_2^2 \ . \tag{3.15}$$

Finally, applying (3.6) to (3.15) yields

$$||\bar{x}_m - x^*||_2^2 \leq \gamma_1^{-1} K(\gamma_2 + \gamma_3 K) \ ||\hat{u}_m - u^*||_2^2 \ . \tag{3.16}$$

We now summarise the above results in the following theorem.

Theorem 3.1

Lte $(\bar{u}_m, \bar{x}_m)$ be the control-state pair prescribed by the CP method over the approximation space C_m for the problem defined by (3.1), (3.2) and the appropriate assumptions. Then

$$||\bar{u}_m - u^*||_2 \leq [\gamma_1^{-1}(\gamma_2 + \gamma_3 K)]^{\frac{1}{2}} \ ||\hat{u}_m - u^*||_2 \tag{3.17}$$

$$||\bar{x}_m - x^*||_2 \leq [\gamma_1^{-1} K(\gamma_2 + \gamma_3 K)]^{\frac{1}{2}} ||\hat{u}_m - u^*||_2 \tag{3.18}$$

and

$$0 \leq J(\bar{u}_m) - J(u^*) \leq (\gamma_2 + \gamma_3 K) \, ||\hat{u}_m - u^*||_2^2 . \tag{3.19}$$

Error Bounds for CP Procedure Over Spline Approximation Spaces

We now specialize the above result to spline approximation spaces and obtain error bounds for the approximate cost, control and state. In this case, it is more appropriate to label spline spaces by the mesh norm h rather than by the dimension m. Assuming that the family of spaces S_h^α (with α fixed) parametrized by h satisfies the condition that $\lim_{h\to 0} S_h^\alpha$ is dense in U, the above error estimates apply with C_m replaced by S_h^α.

Theorem 3.2

Let $(\bar{u}^h, \bar{x}^h)$ be the control-state pair prescribed by the CP method over the approximation space S_h^α for the problem defined by (3.1), (3.2) and the appropriate assumptions. Then

$$||\bar{u}^h - u^*||_2 \leq O(h^\alpha) \tag{3.20}$$

$$||\bar{x}^h - x^*||_2 \leq O(h^\alpha) \tag{3.21}$$

and

$$0 \leq J(\bar{u}^h) - J(u^*) \leq O(h^{2\alpha}). \tag{3.22}$$

Proof

By the approximation properties of splines (see Appendix A) we know that, since $u^* \in \{PC^\alpha[0,T], E^r\}$, there exists $u_s^h \in S_h^\alpha$ such that

$$||u_s^h - u^*||_2 \leq O(h^\alpha). \tag{3.23}$$

The definition of $\hat{u}^h$ is

$$||\hat{u}^h - u^*||_2 \equiv \inf\{||u^h - u^*||_2 | u^h \in S_h^\alpha\},$$

so it is obvious that

$$||\hat{u}^h - u^*||_2 \leq ||u_s^h - u^*||_2 \leq O(h^\alpha). \tag{3.24}$$

Hence, by (3.17) we deduce that

$$||\bar{u}^h - u^*||_2 \leq [\gamma_1^{-1}(\gamma_2 + \gamma_3 K)]^{\frac{1}{2}} ||u^h - u^*||_2 \leq O(h^\alpha) \qquad (3.25)$$

which is the required error bound of (3.20). The remaining error bounds (3.21) and (3.22) follow readily from (3.18) and (3.19).

Numerical Example

In order to test the error bounds of Theorem 3.2, approximate solutions to the following linear quadratic problem were obtained using the CP method.

For the second order system

$$\dot{x}_1(t) = x_2(t) , \quad x_1(0) = 1 ,$$

$$\dot{x}_2(t) = u(t) , \quad x_2(0) = 0 ,$$

find the control $u^*(t)$ that minimizes the cost functional

$$J(u) = \frac{1}{2} \int_0^5 [x_1^2(t) + u^2(t)]dt.$$

This simple pure integrator system was chosen because of the ease with which the exact solution could be obtained; the analytical solution of the above optimal control problem was found to be given by

$$x_1^*(t) = e^{\alpha t}[a_1 \cos(\alpha t) + a_2 \sin(\alpha t)] + e^{-\alpha t}[a_3 \cos(\alpha t) + a_4 \sin(\alpha t)] \qquad (3.26)$$

where $\alpha = 1/\sqrt{2}$, and a_1, a_2, a_3, a_4 are constants given by

$$a_1 = 0.001,689 , \quad a_2 = 0.001,195 , \quad a_3 = 0.998,311 , \quad a_4 = 0.995,428.$$

The other optimal variables can be obtained by differentiating (3.26). Finally, the optimal cost $J(u^*)$ was computed to be

$$J(u^*) = 0.704,7187.$$

Two sets of numerical results were obtained by parametrizing the approximate control firstly as a linear spline, and secondly as a parabolic spline. In both cases a uniform partition of the time interval $[0,5]$ was used, and results were obtained for several values of N, the number of sections in the partition. For each N, the mesh length h = 5/N.

In addition to evaluating the minimum cost $J(\bar{u}_m)$ for each value of N used, the L_2-norms of the errors $\bar{x}_1^h - x_1^*$, $\bar{x}_2^h - x_2^*$ and $\bar{u}^h - u^*$ were also computed where

$$||\bar{x}_1^h - x_1^*||_2 \equiv \{\int_0^5 [\bar{x}_1^h(t) - x_1^*(t)]^2 dt\}^{\frac{1}{2}}, \text{ etc.}$$

In the case where the control is parametrized as a linear spline the integrand in the performance index for this problem is a piecewise polynomial of order 6. Therefore the definite integrals were evaluated by employing the 4-point Gauss-Legendre quadrature formula (which computes exactly the definite integrals of polynomials of degree 7 or less) for each section in the partition. Similarly, for the quadratic spline case the 5-point Gauss-Legendre quadrature formula was employed.

(i) Approximation over S_h^2

The results for this case are presented in Table 3.1

TABLE 3.1 CONVERGENCE HISTORIES

N	$J(\bar{u}^h)$	$J(\bar{u}^h)-J(u^*)$	$\|\|\bar{x}_1^h-x_1^*\|\|_2$	$\|\|\bar{x}_2^h-x_2^*\|\|_2$	$\|\|\bar{u}^h-u^*\|\|_2$
2	0.743 359	0.038 640	0.883E-1	0.129E0	0.264E0
3	0.711 747	0.007 028	0.169E-1	0.357E-1	0.117E0
4	0.706 774	0.002 055	0.478E-2	0.137E-1	0.639E-1
5	0.705 525	0.000 806	0.180E-2	0.661E-2	0.401E-1
6	0.705 098	0.000 379	0.861E-3	0.372E-2	0.275E-1

To see how well the above results fit the convergence rates predicted by Theorem 3.2, we construct the Table 3.2 of convergence rates using Table 3.1. The entity α_N in Table 3.2 is defined by

$$\alpha_N = \log \frac{\left[J(\bar{u}^h) - J(u^*)\right]_{N-1}}{\left[J(\bar{u}^h) - J(u^*)\right]_N} \Big/ \log \left(\frac{N}{N-1}\right) ,$$

and β_N, γ_N and δ_N are similarly defined for $||\bar{x}_1^h - x_1^*||_2$, $||\bar{x}_2^h - x_2^*||_2$ and $||\bar{u}^h - u^*||_2$ respectively.

TABLE 3.2 CONVERGENCE RATES

N	α_N	β_N	γ_N	δ_N
3	4.20	4.09	3.16	1.99
4	4.27	4.38	3.34	2.11
5	4.19	4.37	3.25	2.08
6	4.14	4.06	3.15	2.06

The convergence rates predicted by Theorem 3.2 for the approximation space S_h^2 are

$$||\bar{u}^h - u^*||_2 \leq O(h^2) ,$$

$$||\bar{x}_i^h - x^*||_2 \leq O(h^2) , \quad i = 1,2 ,$$

$$J(\bar{u}^h) - J(u^*) \leq O(h^4) .$$

From Table 3.2 it may be observed that the convergence rates for $J(\bar{u})$ and $\bar{u}$ agree closely with the theoretically predicted rates. However, the observed convergence rates for $\bar{x}_1$ and $\bar{x}_2$ are actually higher than the $O(h^2)$ predicted from theory. This is because the present numerical example is a pure integrator system, for which the order bound for the

state approximation $\bar{x}$ stated in Theorem 3.2 is not optimal. Improved order bounds for this special case are derived in Chapter 5.

(ii) Approximation over S_h^3

The results for this case are presented below in Table 3.3.

TABLE 3.3 CONVERGENCE HISTORIES

N	$J(\bar{u}^h)$	$J(\bar{u}^h)-J(u^*)$	$\|\|\bar{x}_1^h-x_1^*\|\|_2$	$\|\|\bar{x}_2^h-x_2^*\|\|_2$	$\|\|\bar{u}^h-u^*\|\|_2$
2	0.706 2359	0.001 5172	0.919E-2	0.199E-1	0.543E-1
3	0.704 8297	0.000 1110	0.117E-2	0.389E-2	0.148E-1
4	0.704 7428	0.000 0241	0.412E-3	0.157E-2	0.693E-2
5	0.704 7254	0.000 0067	0.135E-3	0.657E-3	0.365E-2
6	0.704 7209	0.000 0022	0.512E-4	0.311E-3	0.212E-2

Table 3.4 of convergence rates was constructed in a similar way to Table 3.2.

TABLE 3.4 CONVERGENCE RATES

N	α_N	β_N	γ_N	δ_N
3	6.45	5.08	4.03	3.20
4	5.31	3.63	3.15	2.64
5	5.74	5.00	3.90	2.87
6	6.11	5.32	4.10	2.98

The convergence rates predicted by Theorem 3.2 for the approximation space S_h^3 are

$$||\bar{u}^h - u^*||_2 \leq O(h^3) ,$$

$$||\bar{x}_i^h - x_i^*||_2 \leq O(h^3) , \quad i = 1,2 ,$$

$$J(\bar{u}^h) - J(u^*) \leq O(h^6) .$$

We observe that the convergence rates for $J(\bar{u})$ and $\bar{u}$ are in good agreement with those observed in Table 3.4. However, we note that the convergence rates for $\bar{x}_1$ and $\bar{x}_2$ in Table 3.4 are better than the $O(h^3)$ predicted from theory. The reason for this behaviour is the same as that given previously for the S_h^2 case.

The optimal control and state profiles are tabulated below in Tables 3.5 and 3.6 together with the CP solutions for N=2 and N=6.

TABLE 3.5 CONTROL PROFILES

t	N=2 $\bar{u}(t)$	N=6 $\bar{u}(t)$	$u^*(t)$
0.0	-0.90080E0	-0.99343E0	-0.99423E0
0.5	-0.43573E0	-0.41137E0	-0.41227E0
1.0	-0.92166E-1	-0.56175E-1	-0.53747E-1
1.5	0.12989E0	0.13327E0	0.13077E0
2.0	0.23044E0	0.19471E0	0.19590E0
2.5	0.20948E0	0.18962E0	0.18931E0
3.0	0.13605E0	0.14738E0	0.14719E0
3.5	0.79212E-1	0.93901E-1	0.94250E-1
4.0	0.38952E-1	0.46347E-1	0.46117E-1
4.5	0.15275E-1	0.12395E-1	0.12483E-1
5.0	0.81797E-2	-0.13557E-3	0.00000E0

TABLE 3.6 STATE TRAJECTORIES

t	N=2 $\bar{x}_1(t)$	N=6 $\bar{x}_1(t)$	$x_1^*(t)$
0.0	0.10000E1	0.10000E1	0.10000E1
0.5	0.90804E0	0.90249E0	0.90250E0
1.0	0.70462E0	0.69732E0	0.69725E0
1.5	0.47563E0	0.47487E0	0.47493E0
2.0	0.27658E0	0.28286E0	0.28283E0
2.5	0.13260E0	0.13822E0	0.13822E0
3.0	0.39902E-1	0.40214E-1	0.40217E-1
3.5	-0.18438E-1	-0.21205E-1	0.21207E-1
4.0	-0.56630E-1	-0.58964E-1	0.58962E-1
4.5	-0.84738E-1	-0.84881E-1	0.84884E-1
5.0	-0.10868E0	-0.10724E0	0.10724E0

3.3 UNCONSTRAINED PROBLEMS - GENERAL CASE

Problem Statement

Consider the general nonlinear optimal control problem described by the state equation

$$\dot{x}(t) = f(x,u,t) \quad , \quad x(0) = x_o \quad , \quad t \in [0,T] \tag{3.27}$$

and the cost functional

$$J(u) = \int_0^T \phi(x,u,t)dt \tag{3.28}$$

where $x(t)$ is an n-dimensional vector belonging to a state space X, $u(t)$ is an r-dimensional vector belonging to the class of admissible controls U, $f(x,u,t)$ is an n-dimensional vector-valued function, and $\phi(x,u,t)$ is a scalar-valued function.

Following Bosarge et al. [2], the following assumptions are made on the problem:

(A1) $f(x,u,t)$ and $\phi(x,u,t)$ are continuously differentiable $\alpha + 1$ times in x and u, and α times in t, where $\alpha \geq 1$.

(A2) The operators $\partial^{i+j} f/(\partial x^i \partial u^j)$ and $\partial^{i+j}\phi/(\partial x^i \partial u^j)$ map bounded neighbourhoods of $X \times U$ into bounded neighbourhoods of $\{L_2[0,T], E^n\}$ and $L_2[0,T]$ respectively, for positive integers i and j where $i+j \leq \alpha+1$.

(A3) The necessary conditions for optimality at (x^*,u^*) are satisfied:

$$\dot{x}^*(t) = f(x^*,u^*,t), \quad x^*(0) = x_o , \tag{3.29}$$

$$\dot{\lambda}^*(t) = -\frac{\partial H}{\partial x}(x^*,u^*,t) , \quad \lambda^*(T) = 0 \tag{3.30}$$

$$\frac{\partial H}{\partial u}(x^*,u^*,\lambda^*,t) = 0 , \tag{3.31}$$

where H is the Hamiltonian defined in the usual fashion,

$$H(x,u,\lambda,t) = \phi(x,u,t) + \langle \lambda, f(x,u,t) \rangle \tag{3.32}$$

(A4) The second variation of the cost functional J is strongly positive at the optimum; i.e.

$$\int_0^T \left\langle H''(x^*,u^*,\lambda^*,t) \begin{bmatrix} \delta u \\ \delta x \end{bmatrix} , \begin{bmatrix} \delta u \\ \delta x \end{bmatrix} \right\rangle dt \geq \sigma_1 ||\delta u||_2^2 , \tag{3.33}$$

where

$$H'' \equiv \begin{bmatrix} \partial^2 H/\partial u^2 & \partial^2 H/\partial u \partial x \\ \partial^2 H/\partial x \partial u & \partial^2 H/\partial x^2 \end{bmatrix} , \qquad \sigma_1 > 0 \ (\sigma_1 \text{ constant}),$$

and

$$\dot{\delta x} = \frac{\partial f}{\partial x}(x^*,u^*,t)\ \delta x + \frac{\partial f}{\partial u}(x^*,u^*,t)\,\delta u , \quad \delta x(0) = 0 \tag{3.34}$$

Assumptions (A1 - (A4) constitute a set of local sufficiency conditions for optimality (see [7]). Note that (A4) is weaker than the

corresponding assumption of Bosarge et.al. [2]. Using these assumptions and standard differential equation arguments, it can be established that $u^* \in \{C^\alpha[0,T], E^r\}$ and $x^* \in \{C^\alpha[0,T], E^n\}$. Thus, as in the linear quadratic case, we can choose $U = \{W_2^\alpha[0,T], E^r\}$ and $X = \{W_2^\alpha[0,T], E^n\}$.

Remarks

Assumptions (A2) and (A4) imply that there exist bounded convex neighbourhoods $N(u^*) \subset U$ of u^* and $N(x^*) \subset X$ of x^* such that

(i) f satisfies a Lipschitz condition with respect to x and u,

$$\left| f(x^*+\Delta x,\ u^*+\Delta u, t) - f(x^*,u^*,t) \right| \leq L\left(\left|\Delta x(t)\right| + \left|\Delta u(t)\right| \right) \tag{3.35}$$

for $x^*+\Delta x \in N(x^*)$, $u^*+\Delta u \in N(u^*)$, where L is a positive constant and $|.|$ denotes the usual Euclidean norm (i.e. $|\Delta x(t)| = \langle \Delta x(t), \Delta x(t) \rangle^{\frac{1}{2}}$ etc.).

(ii)
$$\int_0^T \left\langle H''(\tilde{x},\tilde{u},\lambda,t) \begin{bmatrix} \Delta u \\ \Delta x \end{bmatrix}, \begin{bmatrix} \Delta u \\ \Delta x \end{bmatrix} \right\rangle dt \geq 2\sigma \left|\left| \delta u \right|\right|_2^2 , \tag{3.36}$$

where $\tilde{u} \in N(u^*)$, $\tilde{x} \in N(x^*)$, Δx is given by

$$\Delta\dot{x}(t) = f(x^*+\Delta x,\ u^*+\Delta u, t) - f(x^*,u^*,t), \quad \Delta x(0) = 0 , \tag{3.37}$$

and σ is a positive constant.

We shall use conditions (i) and (ii) in the sequel.

Perturbation About the Optimal Control

Consider a perturbation Δu about the optimal control u^* which produces a perturbation Δx about x^* governed by the differential equation

$$\Delta\dot{x}(t) = f(x,u,t) - f(x^*,u^*,t) , \quad \Delta x(0) = 0 \tag{3.38}$$

It is assumed that $u \equiv u^* + \Delta u \in N(u^*)$ and $x \equiv x^* + \Delta x \in N(x^*)$. We can write (3.38) as

$$\Delta x(t) = \int_0^t \left[f(x,u,\tau) - f(x^*,u^*,\tau) \right] d\tau \tag{3.39}$$

which means that

$$|\Delta x(t)| \leq \int_0^t |f(x,u,\tau) - f(x^*,u^*,\tau)|\,d\tau. \tag{3.40}$$

And using the Lipschitz condition (3.35) we have

$$|\Delta x(t)| \leq L\int_0^t (|\Delta x(\tau)| + |\Delta u(\tau)|)\,d\tau \tag{3.41}$$

Applying the Gronwall Inequality [6] to (3.41), we then have

$$||\Delta x||_2^2 \leq K\,||\Delta u||_2^2 \tag{3.42}$$

for some constant $K > 0$.

We next consider the effect of the perturbation Δu on the cost J. Expanding J in a Taylor series (see [8]) about u^*, we have

$$J(u) = J(u^*) + \int_0^T \left\langle \frac{\partial H}{\partial u}(x^*,u^*,\lambda^*,t), \Delta u \right\rangle dt$$

$$+ \int_0^1 (1-\zeta)\int_0^T \left\langle H''(\tilde{x},\tilde{u},\lambda^*,t) \begin{bmatrix}\Delta u\\ \Delta x\end{bmatrix}, \begin{bmatrix}\Delta u\\ \Delta x\end{bmatrix} \right\rangle dt\, d\zeta, \tag{3.43}$$

where $\tilde{u} = \zeta u + (1-\zeta)u^*$ and $\tilde{x} = \zeta x + (1-\zeta)x^*$.

By assumption (A3) $\partial H/\partial u$ vanishes along an optimizing arc, so that

$$J(u) - J(u^*) = \int_0^1 (1-\zeta)\int_0^T \left\langle H''(x,u,\lambda,t) \begin{bmatrix}\Delta u\\ \Delta x\end{bmatrix}, \begin{bmatrix}\Delta u\\ \Delta x\end{bmatrix} \right\rangle dt\, d\zeta. \tag{3.44}$$

Using (3.36), we can obtain a lower bound for $J(u) - J(u^*)$. An upper bound can be obtained by applying the Cauchy-Schwarz inequality to (3.44). Thus we have

$$\sigma||\Delta u||_2^2 \leq J(u) - J(u^*) \leq \rho\left[||\Delta x||_2^2 + ||\Delta u||_2^2\right] \tag{3.45}$$

for some constant $\rho > 0$. Combining (3.42) and (3.45) we finally have

$$\sigma||\Delta u||_2^2 \leq J(u) - J(u^*) \leq \rho(1+K)\,||\Delta u||_2^2 . \tag{3.46}$$

Error Estimates

Let the approximation space C_m and the entities $\bar{u}_m$, $\hat{u}_m$ be defined as before in the linear quadratic case. We are interested in the convergence properties of the CP procedure as $m \to \infty$. By our assumption that

$C = \lim_{m\to\infty} C_m$ is dense in U, it follows that the best L_2-approximation $\hat{u}_m \in N(u^*)$ for m sufficiently large. In view of (3.46) it is also clear that the CP solution $\bar{u}_m \in N(u^*)$ for m sufficiently large.

By definition of $\bar{u}_m$, we have

$$0 \leq J(\bar{u}_m) - J(u^*) \leq J(\hat{u}_m) - J(u^*) \ , \tag{3.47}$$

and using the upper bound in (3.46) for $u = \hat{u}_m$, we obtain

$$0 \leq J(\bar{u}_m) - J(u^*) \leq \rho(1+K)\,||\hat{u}_m - u^*||_2^2. \tag{3.48}$$

Next, combining (3.48) with the lower bound in (3.46) for $u = \bar{u}_m$, we have

$$||\bar{u}_m - u^*||_2^2 \leq \sigma^{-1}\rho(1+K)\,||\hat{u}_m - u^*||_2^2. \tag{3.49}$$

Finally, using (3.42) we obtain that

$$||\bar{x}_m - x^*||_2^2 \leq \sigma^{-1}\rho K(1+K)\,||\hat{u}_m - u^*||_2^2. \tag{3.50}$$

These results are summarised in the following theorem.

Theorem 3.3

Let $(\bar{u}_m, \bar{x}_m)$ be the control-state pair prescribed by the CP procedure over C_m for the problem defined by (3.27), (3.28) and assumptions (A1) - (A4). Then for $m\to\infty$,

$$||\bar{u}_m - u^*||_2 \leq \left[\sigma^{-1}\rho(1+K)\right]^{\frac{1}{2}}||\hat{u}_m - u^*||_2 \tag{3.51}$$

$$||\bar{x}_m - x^*||_2 \leq \left[\sigma^{-1}\rho K(1+K)\right]^{\frac{1}{2}}||\hat{u}_m - u^*||_2 \tag{3.52}$$

and

$$0 \leq J(\bar{u}_m) - J(u^*) \leq \rho(1+K)\,||\hat{u}_m - u^*||_2^2. \tag{3.53}$$

Error Bounds Over Spline Spaces

We now specialize the above result to the spline approximation space S_h^α. Employing arguments outlined for Theorem 3.2, the following result may be readily deduced from Theorem 3.3.

Theorem 3.4

Let $(\bar{u}^h, \bar{x}^h)$ be the control-state pair prescribed by the CP procedure over S_h^α for the problem defined by (3.27), (3.28) and assumptions (A1) - (A4). Then

$$||\bar{u}^h - u^*||_2 \leq O(h^\alpha) \tag{3.54}$$

$$||\bar{x}^h - x^*||_2 \leq O(h^\alpha) \tag{3.55}$$

and

$$0 \leq J(\bar{u}^h) - J(u^*) \leq O(h^{2\alpha}). \tag{3.56}$$

3.4 PROBLEMS WITH FIXED TERMINAL STATE

In the previous section an error analysis of the CP method as applied to unconstrained control problems was carried out. The analysis is extended in this section to problems with specified state at the final time. For simplicity, we will only consider the stationary linear regulator problem. At the same time there is less emphasis on mathematical rigor in this section than in the previous one, and the following exposition will be found to contain some rather heuristic arguments. However, it is believed that the main ideas and conclusions of this section should not be affected in any way.

Problem Statement

Consider the following n-dimensional constant linear system with a scalar control variable

$$\dot{x}(t) = A\, x(t) + b\, u(t) \tag{3.57}$$

with the specified end points

$$x(0) = x_o \ , \quad x(T) = x_f \tag{3.58}$$

and the quadratic cost functional

$$J(u) = \int_0^T \left[\langle x(t), Qx(t) \rangle + \langle u(t), Ru(t) \rangle \right] dt. \tag{3.59}$$

Here A is an n × n constant matrix, b is a constant n-vector, Q is an n × n constant symmetric positive semi-definite matrix and R is simply a positive constant.

By standard differential equation arguments, it can be easily shown that the optimal control $u^*(t)$ for the above problem is infinitely smooth.

Spline Controllability

Consider the following constant linear system

$$\dot{x}(t) = Ax(t) + bu(t) \tag{3.60}$$

with the specified end points

$$x(0) = 0 \ , \quad x(T) = \beta \ , \tag{3.61}$$

where β is an n-dimensional vector.

Let S_h^α be the space of splines of arbitrary order α-1 defined on a uniform mesh of $[0,T]$. Assume that the above system (3.60) is completely controllable. It was pointed out in Chapter 2 that provided the mesh h is sufficiently small, the system can be transferred from its initial state $x(0) = 0$ to its final state $x(T) = \beta$ by a control belonging to S_h^α. We now wish to construct such a control u_c^h whose "size" is proportional to the "size" of β.

Suppose that $\omega_i^h \in S_h^\alpha$ transfers the system from the initial state $x(0) = 0$ to the final state $x(T) = e_i$ for $i = 1,2,\ldots,n$, where e_i denotes the n-vector whose i^{th}-component is one and whose remaining components are zero. Then, by the principle of superposition, the control

$$u_c^h = \sum_{i=1}^{n} \beta_i \omega_i^h \tag{3.62}$$

belongs to S_h^α and transfers the system from $x(0) = 0$ to the final state

$$x(T) = \sum_{i=1}^{n} \beta_i e_i = \beta \ . \tag{3.63}$$

The size of the control u_c^h as measured by the L_2-norm is

$$||u_c^h||_2 \leq \sum_{i=1}^{n} |\beta_i| \ ||\omega_i^h||_2 \ . \tag{3.64}$$

Intuitively, it would appear that $||\omega_i^h||_2$ is bounded for all h. For instance, if h_1 is an integral multiple of h_2 we can choose the functions $\omega_i^{h_1}$ and $\omega_i^{h_2}$ so that $||\omega_i^{h_1}||_2 \leq ||\omega_i^{h_2}||$. Let us now accept that $||\omega_i^h||_2$ is bounded for all h and for i = 1,2,...,n; that is, $||\omega_i^h||_2 \leq K$ for some positive constant K. It follows from (3.64) that

$$||u_c^h||_2 \leq K \sum_{i=1}^{n} |\beta_i| \ . \tag{3.65}$$

Convergence Result

Ignoring the terminal constraint for the moment, we know from the theory of spline approximation that there exists a family of controls $u^h \in S_h^\alpha$ parametrized by h, such that

$$||u_s^h - u^*||_2 \leq O(h^\alpha). \tag{3.66}$$

It follows from the Gronwall Inequality [6] that the state trajectory generated by the control u_s^h with the initial condition $x(0) = x_o$ satisfies the inequality

$$||x_s^h - x^*||_2 \leq O(h^\alpha). \tag{3.67}$$

However, in general, the trajectory x_s^h will not satisfy the terminal constraint $x(T) = x_f$. The violation of the terminal constraint is given by

$$x_s^h(T) - x_f = \int_0^T \Phi(T-\tau) b \left[u_s^h(\tau) - u^*(\tau)\right] d\tau \ , \tag{3.68}$$

where $\Phi(T-\tau)$ is the transition matrix for the system. Let $\zeta(\tau) \equiv \Phi(T-\tau)b$; considering the i-th component of the equation (3.68), we have

$$\begin{aligned} \left| \left[x_s^h(T) - x_f\right]_i \right| &= \left| \int_0^T \zeta_i(\tau)\left[u_s^h(\tau) - u^*(\tau)\right] d\tau \right| \\ &\leq ||\zeta_i||_2 \, ||u_s^h - u^*||_2 \\ &\leq O(h^\alpha) \qquad \text{(from (3.66))}. \end{aligned} \tag{3.69}$$

Let $\beta_i = \left[x_s^h(T) - x_f\right]_i$, and let u_c^h be defined by (3.62). It then follows from (3.65) and (3.69) that

$$||u_c^h||_2 \leq O(h^\alpha). \tag{3.70}$$

Moreover, it is easy to see that the control

$$\tilde{u}^h = u_s^h + u_c^h \tag{3.71}$$

transfers the system from the initial state $x(0) = x_o$ to the final state $x(T) = x_f$. And it is also clear from (3.66) and (3.70) that

$$||\tilde{u}^h - u^*||_2 \leq O(h^\alpha). \tag{3.72}$$

Thus we have constructed a family of admissible controls $\tilde{u}^h \in S_h^\alpha$ satisfying the terminal constraint $x(T) = x_f$ and approximating u^* to order α. We are now ready to proceed with the derivation of error bounds.

Let U_f be the subset of the admissible space of controls U which satisfies the terminal constraint $x(T) = x_f$. The CP method seeks the control $\bar{u}^h \in S_h^\alpha \cap U_f$ such that

$$J(\bar{u}^h) = \inf\{J(u^h) \,|\, u^h \in S_h^\alpha \cap U_f\}. \tag{3.73}$$

It is then obvious that

$$J(\bar{u}^h) \leq J(\tilde{u}^h).$$

Now, it can easily be shown that the identity (3.7) is also valid for linear quadratic problems with terminal constraints. Hence it follows that

$$||\bar{x}^h - x^*||_Q^2 + ||\bar{u}^h - u^*||_R^2 \leq ||\tilde{x}^h - x^*||_Q^2 + ||\tilde{u}^h - u^*||_R^2 . \quad (3.74)$$

And using (3.3), (3.4) and (3.6) we can conclude from (3.74) that

$$||\bar{u}^h - u^*||_R^2 \leq (\gamma_3 K+\gamma_2)||\tilde{u}^h - u^*||_2^2 \quad (3.75)$$

Applying (3.3) to (3.75), we obtain

$$||\bar{u}^h - u^*||_2 \leq [\gamma_1^{-1}(\gamma_3 K+\gamma_2)]^{\frac{1}{2}}||\tilde{u}^h - u^*||_2 \quad (3.76)$$

$$\leq O(h^\alpha).$$

Similarly, we can also show that

$$||\bar{x}^h - x^*||_2 \leq O(h^\alpha) \quad (3.77)$$

and

$$J(\bar{u}^h) - J(u^*) \leq O(h^{2\alpha}). \quad (3.78)$$

Hence we see that the error bounds for the problem with terminal constraint are as good as those which were obtained for the unconstrained problem.

Numerical Example

To check the error bounds derived above, the CP method was applied to the following linear quadratic problem.

For the second order system

$$\dot{x}_1(t) = x_2(t) ,$$

$$\dot{x}_2(t) = u(t) ,$$

with the boundary conditions

$$x_1(0) = 1 \qquad x_1(1) = 0$$

$$x_2(0) = 0 \qquad x_2(1) = 0 \ ,$$

find the control $u^*(t)$ that minimizes the cost functional

$$J(u) = \frac{1}{2} \int_0^1 (x_1^2(t) + .01\ u^2(t))\ dt \ .$$

The analytical solution to the above problem can be shown to be

$$x_1^*(t) = e^{\alpha t}\left[a_1 \cos(\alpha t) + a_2 \sin(\alpha t)\right] + e^{-\alpha t}\left[a_3 \cos(\alpha t) + a_4 \sin(\alpha t)\right] \tag{3.79}$$

where $\alpha = \sqrt{5}$ and a_1, a_2, a_3, a_4 are constants given by

$$a_1 = -.015{,}109 \ , \quad a_2 = -.014{,}693 \ , \quad a_3 = 1.015{,}109,$$

$$a_4 = 1.044{,}911 \ .$$

The other optimal variables x_2^* and u^* can be obtained by differentiating (3.79). Finally, the optimal cost J^* was computed to be

$$J^* = 0.230{,}364.$$

Approximate solutions to the problem were obtained by parametrizing the control as a linear spline over a uniform partition of the time interval $[0,1]$. Results were obtained for several values of N, the number of sections in the partition. These numerical results are summarised in the following Table 3.7.

TABLE 3.7 CONVERGENCE HISTORIES

N	$J(\bar{u}^h)$	$J(\bar{u}^h)-J(u^*)$	$\|\|\bar{x}_1^h-x_1^*\|\|_2$	$\|\|\bar{x}_2^h-x_2^*\|\|_2$	$\|\|\bar{u}^h-u^*\|\|_2$
2	0.232 077	0.001 713	0.779E-2	0.538E-1	0.580E0
3	0.230 664	0.000 300	0.131E-2	0.137E-1	0.245E0
4	0.230 455	0.000 091	0.398E-3	0.547E-2	0.135E0
5	0.230 400	0.000 017	0.159E-3	0.272E-2	0.853E-1

From the above Table, we construct a Table of convergence rates as in the previous section. The quantities α_N, β_N, γ_N, δ_N are defined as for Table 3.2.

TABLE 3.8 CONVERGENCE RATES

N	α_N	β_N	γ_N	δ_N
3	4.30	4.40	3.37	2.13
4	4.15	4.14	3.20	2.08
5	4.10	4.11	3.13	2.04

The convergence rates in Table 3.8 are in good agreement with those obtained from the preceding error analysis. These are given by equations (3.76), (3.77) and (3.78), which, for this particular case $\bar{u}^h \in S_h^2$, are as follows:

$$J(\bar{u}^h) - J(u^*) \leq O(h^4) , \tag{3.80}$$

$$||\bar{x}_i^h - x_i^*||_2 \leq O(h^2) , \quad i = 1,2, \tag{3.81}$$

$$||\bar{u}^h - u^*||_2 \leq O(h^2). \tag{3.82}$$

Again, as in section 3.2, we notice that the convergence rates indicated by the computation results are better than those predicted by (3.81) for the state variables.

3.5 CONCLUSIONS

Local convergence of CP solutions to the optimal control problem has been studied. Error estimates for the control, state and cost functional have been derived for the CP approximation over arbitrary finite-dimensional spaces. These error estimates have been derived in the mean square norm, initially for the unconstrained linear quadratic

problem and later for the general unconstrained problem.

Explicit order bounds have also been obtained for the CP approximation over spline spaces. It was found that by restricting the control to the space S_h^α of splines of order $\alpha-1$, the CP procedure would deliver approximations to the control, state and cost with the orders of accuracy $O(h^\alpha)$, $O(h^\alpha)$ and $O(h^{2\alpha})$ respectively. In comparison, by restricting the co-state to the space S_h^α, the Ritz-Trefftz procedure would deliver approximations to the control, state and cost with the orders of accuracy $O(h^{\alpha-1})$, $O(h^{\alpha-1})$ and $O(h^{2\alpha-2})$ respectively (see [1]).

It has also been demonstrated that the error bounds obtained for the unconstrained problem remain valid for the problem with terminal constraints. Finally, results from numerical experiments have been used to check the error bounds which had been derived theoretically. It was found that the error bounds observed from the numerical results agreed very well with those predicted from theory in the case of the cost functional convergence and the control convergence. However, the state variables were observed to converge at higher rates than were predicted. This is because we used examples involving pure integrator systems. In Chapter 5 it is shown that higher order bounds for the state variables than those predicted by Theorem 3.2 apply for pure integrator systems.

REFERENCES

[1] W. E. Bosarge, Jr. and O. G. Johnson: Error Bounds of High Order Accuracy for the State Regulator Problem via Piecewise Polynomial Approximations, SIAM J. Control, Vol.9 (1971) pp 15-28.

[2] W. E. Bosarge, Jr., O. G. Johnson, R. S. McKnight and W. P. Timlake: The Ritz-Galerkin Procedure for Nonlinear Control Problems, SIAM J. Numer. Anal., Vol.10 (1973) pp 94-111.

[3] M. H. Schultz: Spline Analysis, Prentice-Hall Inc., N.J. (1973).

[4] E. A. Coddington and N. Levinson: Theory of Ordinary Differential Equations, McGraw-Hill, N.Y. (1955).

[5] P. J. Davis: Interpolation and Approximation, Blaisdell, N.Y. (1955).

[6] R. Bellman: Stability Theory of Differential Equations, McGraw-Hill, N.Y. (1952).

[7] I. M. Gelfand and S. V. Fomin: Calculus of Variations, Prentice-Hall Inc., N.J. (1963).

[8] J. Dieudonné: Foundations of Modern Analysis, Academic Press, N.Y. (1960).

[9] R. E. Kalman, Y. C. Ho and K. S. Narendra: Controllability of Linear Dynamical Systems, in Contributions to Differential Equations, Vol.1, Interscience Publishers (1963).

CHAPTER 4

THE STATE PARAMETRIZATION PROCEDURE

4.1 INTRODUCTION

In the preceding chapters we examined the CP procedure for solving optimal control problems. It was pointed out by Mehra and Davis [1] that although it is natural to treat the control variables as the independent variables and the state variables as the dependent ones in the system equation, it is not essential to do so. In certain situations, particularly those involving terminal constraints or state variable inequality constraints, it is more convenient to treat some of the state variables as independent variables. This is the basis for the generalized gradient (GRG) approach for optimal control problems presented in [1].

In this chapter we present a new method for solving optimal control problems by combining the GRG approach with the parametrization approach. We shall refer to this method as the state parametrization (SP) procedure. The SP procedure is formulated for general optimal control problems in section 4.2. Although in theory the dependent variables can always be solved for the independent variables through the state equation, in practice this could be a cumbersome task. However there is a wide class of systems to which the SP procedure is particularly well suited. For these systems the dependent variables can be determined in a relatively simple manner without needing to integrate the state equation, thereby increasing the computational efficiency of the procedure. In section 4.3 the SP procedure is formulated for a suitable class of optimal control problems. Then in section 4.4 the SP procedure employing splines is applied to a class of control problems with linear terminal constraints. Numerical results for two sample problems are also presented.

4.2 THE SP PROCEDURE FOR GENERAL CONTROL PROBLEMS

Problem Statement

Consider the general optimal control problem described by the state equation

$$\dot{x}(t) = f(x,u,t) \ , \quad x(0) = x_o \tag{4.1}$$

and the cost functional

$$J = \int_0^T \phi(x,u,t)\ dt \ , \tag{4.2}$$

where x is an n-dimensional state vector and u is an r-dimensional control vector.

Our objective is to find the control u* that minimizes J subject to (4.1).

Method of Approach

The initial step in the SP procedure is to select a set of independent variables from the state variables $x_1,x_2,\ldots,x_n$. Suppose now that $r < n$; without loss of generality, we can assume that the chosen independent variables are $x_1,x_2,\ldots x_r$. It is assumed that the remaining state variables $x_{r+1}, x_{r+2},\ldots x_n$ and the control variables $u_1,u_2,\ldots u_r$ can be uniquely determined in terms of the independent variables through the state equation (4.1).

The next step in the procedure consists of adopting a specific parametrization for the independent variables,

$$x_i(t) = F_i(q_i,t) \ , \quad i = 1,2,\ldots,r \tag{4.3}$$

where each F_i is a known function of the m_i-dimensional parameter vector q_i. It is assumed that the form of the parametrization (4.3) is consistent with the initial conditions of the problem. Let $m \equiv m_1 + m_2 + \ldots + m_r$

and define the m-dimensional parameter vector q by $q^T \equiv [q_1^T | q_2^T \ldots | q_r^T]$. Having parametrized the independent variables by q, the dependent variables can be solved for q through the state equation (4.1) and the cost functional J in (4.2) reduces to a cost function $\tilde{J}(q)$. Hence the original control problem becomes one of minimizing $\tilde{J}(q)$ with respect to the parameter vector q.

As is the case for the CP procedure, it would in general be expedient to perform the optimization of q via a gradient method. The required gradient expressions may be obtained as follows.

Let the Hamiltonian $H(x,u,\lambda,t)$ be defined in the usual fashion,

$$H(x,u,\lambda,t) \equiv \phi(x,u,t) + \langle \lambda, f(x,u,t) \rangle . \tag{4.4}$$

The first variation in the cost J is given by (see [1])

$$\delta J = \langle g_x(T), \delta x(T) \rangle + \int_0^T \left[\langle g_u, \delta u \rangle + \langle g_x, \delta x \rangle \right] dt \tag{4.5}$$

where

$$g_x(T) \equiv -\lambda(T), \quad g_x \equiv \frac{\partial H}{\partial x} + \dot{\lambda}, \quad g_u \equiv \frac{\partial H}{\partial u} .$$

In this case where $\{x_1, x_2, \ldots x_r\}$ are chosen to be the independent variables, the gradients of J wrt the dependent variables $\{x_{r+1}, x_{r+2}, \ldots, x_n, u_1, u_2, \ldots, u_r\}$ are set to zero:

$$g_{x_i}(T) = 0 \quad i = r+1, \ldots, n \tag{4.6}$$

$$g_{x_i} = 0 \quad i = r+1, \ldots, n \tag{4.7}$$

$$g_{u_i} = 0 \quad i = 1,2, \ldots, r \tag{4.8}$$

and the expression (4.5) reduces to

$$\delta J = \sum_{i=1}^{r} \left[g_{x_i}(T) \delta x_i(T) + \int_0^T g_{x_i} \delta x_i \, dt \right] . \tag{4.9}$$

Hence for the parametrization (4.3) we have

$$\frac{\partial \tilde{J}}{\partial q_i} = g_{x_i}(T)\ \frac{\partial F_i}{\partial q_i}\ (q_i,T) + \int_0^T g_{x_i}\ \frac{\partial F_i}{\partial q_i}\ (q_i,t)\ dt\ , \qquad (4.10)$$

$$i = 1,2,\ldots,r\ .$$

Summary of SP Procedure

Step 1: Select the set of independent variables $\{x_1,x_2,\ldots,x_r\}$.

Step 2: Specify the parametrization $x_i = F_i(q_i,t)$, $i = 1,2,\ldots,r$.

Step 3: Set nominal value of q.

Step 4: Determine $x_1,x_2,\ldots,x_r$ from (4.3), and determine the dependent variables $\{x_{r+1},\ldots,x_n,\ u_1,\ldots,u_r\}$ from the state equation (4.1).

Step 5: Solve for $\{\lambda_1,\lambda_2,\ldots,\lambda_n\}$ using the equations:

$$\dot{\lambda}_i + \frac{\partial H}{\partial x_i} = 0\ , \quad \lambda_i(T) = 0\ , \quad i = r+1,\ldots,n$$

$$\frac{\partial H}{\partial u_i} = 0\ , \qquad i = 1,2,\ldots,r$$

Step 6: Evaluate the cost $\tilde{J}(q)$ and the gradients $\partial\tilde{J}/\partial q$ using (4.10).

Step 7: Locate the minimum of $\tilde{J}$ in search direction.

Step 8: Apply appropriate test for convergence. If test fails, update the search direction using a suitable algorithm and return to Step 4.

4.3 THE SP PROCEDURE FOR A CLASS OF OPTIMAL CONTROL PROBLEMS

One of the chief drawbacks of the SP procedure is that the determination of the dependent variables in terms of the independent ones from the state equation (4.1) is in general a cumbersome task. Fortunately there is a wide class of problems for which the above mentioned drawback of the SP procedure does not exist. For these problems the dependent variables can be determined explicitly in a straightforward manner and without having to integrate the state equation numerically. In this section we identify this class of problems and modify the basic SP procedure for these problems.

Problem Statement

We first consider the modified SP procedure for single-input systems. The extension to multivariable systems is considered later on in this section. The class of problems that we are interested in are those involving systems of the phase variable form

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) \\ &\vdots \\ \dot{x}_{n-1}(t) &= x_n(t) \\ \dot{x}_n(t) &= f(x,u,t) \ , \end{aligned} \tag{4.11}$$

where $x = [x_1, x_2, \ldots, x_n]^T$ and f is a scalar-valued function. We assume that the cost functional is given by (4.2), and that the following constraints are imposed at the initial and final times:

$$\zeta[x(0)] = 0 \ , \qquad \eta[x(T)] = 0 \tag{4.12}$$

where ζ and η are vector-valued functions of dimensions ℓ_1 and ℓ_2 respectively. Finally, we assume that the control term u can be solved for x and $\dot{x}_n$ from (4.11):

$$u(t) = g(x, \dot{x}_n, t). \tag{4.13}$$

Remarks

When f is linear in x and u, the system (4.11) reduces to the phase variable canonical form for linear systems. It is well known (see [2]) that a constant linear system can always be transformed into its phase variable canonical form provided that the system is controllable. This is also true for time-varying linear systems under the more restrictive condition of uniform controllability (see [3]).

Method of Approach

The state variable x_1 is chosen as the independent variable and parametrized

$$x_1(t) = F(q,t) , \tag{4.14}$$

where q is an m-dimensional parameter vector. The remaining state variables $x_2,\ldots,x_n$ can be found by successive differentiations of the known function F(q,t) with respect to t, and the control u can be found from (4.13).

From (4.11) and (4.14), we see that

$$x(0) = \left[F(q,0) , \dot{F}(q,0),\ldots,F^{(n-1)}(q,0)\right]^T, \tag{4.15}$$

so we can regard $\zeta[x(0)]$ as some function $\tilde{\zeta}(q)$ of q. Similarly, $\eta[x(T)]$ can also be regarded as some function $\tilde{\eta}(q)$ of q. Hence, the terminal constraints (4.12) can be regarded as $\ell_1 + \ell_2$ side conditions on the parameter vector q.

Of course, it is implicitly assumed here that the form of the parametrization (4.14) is such that the algebraic equations $\tilde{\zeta}(q) = 0$, $\tilde{\eta}(q) = 0$ do possess a solution.

Our original optimal control problem then becomes an optimization problem in q subject to the equality constraints $\tilde{\zeta}(q) = 0$, $\tilde{\eta}(q) = 0$. Problems of this type can be handled using numerical techniques like the gradient projection algorithm of Rosen [9].

Evaluation of Gradients

For the class of problems under consideration, the computation of the required gradients of $\tilde{J}$ with respect to q can be performed in a simpler manner to that described in section 4.2 for general problems. It is assumed here that g and ϕ are differentiable functions of their arguments.

From (4.2), we have that

$$\frac{\partial \tilde{J}}{\partial q_i} = \int_0^T \left[< \frac{\partial \phi}{\partial x} , \frac{\partial x}{\partial q_i} > + \frac{\partial \phi}{\partial u} \frac{\partial u}{\partial q_i} \right] dt , \quad i = 1,\ldots,m. \tag{4.16}$$

It is clear from (4.11) and (4.14) that

$$\frac{\partial x_j}{\partial q_i} = \frac{\partial F^{(j-1)}(q,t)}{\partial q_i} , \quad j = 1,\ldots,n. \tag{4.17}$$

and from (4.13) we obtain

$$\frac{\partial u}{\partial q_i} = < \frac{\partial g}{\partial x} , \frac{\partial x}{\partial q_i} > + \frac{\partial g}{\partial \dot{x}_n} \frac{\partial \dot{x}_n}{\partial q_i} \tag{4.18}$$

where $\frac{\partial x}{\partial q_i}$ is given by (4.17) and $\frac{\partial \dot{x}_n}{\partial q_i}$ is given by

$$\frac{\partial \dot{x}_n}{\partial q_i} = \frac{\partial F^{(n)}(q,t)}{\partial q_i} , \quad i = 1,\ldots,n. \tag{4.19}$$

To implement the gradient projection algorithm we also need to evaluate the gradients of $\tilde{\zeta}$ and $\tilde{\eta}$ with respect to q. It is easily seen that

$$\frac{\partial \tilde{\zeta}(q)}{\partial q_i} = < \frac{\partial \zeta[x(0)]}{\partial x(0)} , \frac{\partial x(0)}{\partial q_i} > , \quad i = 1,\ldots,m \tag{4.20}$$

where $\frac{\partial x(0)}{\partial q_i}$ is computed using (4.15). The evaluation of $\frac{\partial \tilde{\eta}(q)}{\partial q}$ is similar.

Summary of Modified SP Procedure

Step 1: Specify the parametrization (4.14).

Step 2: Determine a feasible value of q which satisfies the constraints $\tilde{\zeta}(q) = 0$, $\tilde{\eta}(q) = 0$. This can be done by minimizing the error function

$$\langle\tilde{\zeta},\tilde{\zeta}\rangle + \langle\tilde{\eta},\tilde{\eta}\rangle \ .$$

Step 3: Evaluate the gradients of $\tilde{\zeta}$ and $\tilde{\eta}$ with respect to q.

Step 4: Compute the projection matrix using the gradients of $\tilde{\zeta}$ and $\tilde{\eta}$.

Step 5: Evaluate the gradient of $\tilde{J}$ using (4.16).

Step 6: Multiply the projection matrix and the gradient vector to find the search direction and then locate the minimum of $\tilde{J}$ in this direction.

Step 7: Apply an appropriate test for convergence. If the test fails, update the projection matrix using the gradient projection algorithm and return to Step 3.

Remarks

We note that whereas Mehra and Davis recommended keeping the control variable as the independent variable in the GRG procedure whenever possible to avoid numerical differentiation of x(t) which can lead to large discontinuities in the control u(t), this difficulty is not encountered when applying the SP procedure to the class of problems considered above because only explicit differentiations of known functions are involved.

Extension to a Class of Multivariable Systems

The class of multivariable systems involving an r-dimensional control vector which is of interest here are those composed of a coupled set of r single-input subsystems each of which is in the standard phase variable

form. The i^{th} subsystem of dimension n_i is of the form

$$\begin{aligned}
\dot{x}_{m_i+1} &= x_{m_i+2} \\
\dot{x}_{m_i+2} &= x_{m_i+3} \\
&\vdots \\
\dot{x}_{m_i+n_i} &= f_i(x,u_i t) \quad , \quad i = 1,2,\ldots,r
\end{aligned} \tag{4.21}$$

where $m_1 \equiv 0$ and $m_i \equiv n_1+\ldots+n_{i-1}$ for $i \geq 2$.

In this case, the state variables $x_{m_i+1}(i=1,\ldots,r)$ are chosen as the independent variables and parametrized. The remaining state variables can be obtained by successive differentiations of the independent variables, while the controls $u_i(i=1,\ldots,r)$ can be determined from (4.21). The evaluation of gradients of the cost with respect to the parameters may be performed in a manner similar to that for single-input systems.

Remarks

When each f_i is linear in x and u_i, the above system reduces to the phase variable canonical from for linear multivariable systems. Luenberger [4] has shown that every constant linear multivariable system which is controllable can be reduced to this phase variable canonical form via non-singular linear transformations of the state and control vectors. Hence the state parametrization procedure developed in this section is applicable to any controllable linear multivariable system.

4.4 PROBLEMS WITH LINEAR CONSTRAINTS

Consider now the optimal control problem described by the equations (4.2), (4.11) and (4.12). In practice, the constraints on the initial state and final state are almost always linear; that is, ζ and η are linear functions. In this section we consider the SP procedure for problems

with linear constraints at the initial and final time.

Suppose that we adopt a linear parametrization for the independent state variable x_1 :

$$x_1(t) = \sum_{i=1}^{m} q_i \psi_i(t) \quad . \tag{4.22}$$

It is easy to see that through the parametrization (4.22), the linear constraints $\zeta[x(0)] = 0$ and $\eta[x(T)] = 0$ reduce to linear algebraic constraints $\tilde{\zeta}(q)$ and $\tilde{\eta}(q)$ in q irrespective of whether f(x,u,t) is linear or nonlinear in x and u. In this case, we can take advantage of the linearity of $\tilde{\zeta}$ and $\tilde{\eta}$ and perform the optimization of $\tilde{J}$ using the quadratically-convergent algorithm of Goldfarb-Lapidus [5].

We note that the above feature is not shared by the CP procedure. A linear parametrization of the control

$$u(t) = \sum_{i=1}^{m} q_i \psi_i(t) \quad , \tag{4.23}$$

does not imply that the linear constraint functions $\zeta[x(0)]$ and $\eta[x(T)]$ will reduce to linear functions $\tilde{\zeta}(q)$ and $\tilde{\eta}(q)$, unless f is also linear.

Numerical Examples

The SP procedure was used to obtain approximate solutions to two sample nonlinear problems. For both problems we adopted a cubic spline parametrization of x_1 over a uniform partition of the time interval concerned:

$$x_1(t) = \sum_{i=1}^{N+3} q_i \psi_i(t) \quad , \tag{4.24}$$

where N is the number of sections in the uniform partition and $\psi_1, \ldots, \psi_{N+3}$ are cubic B-splines.

Example 1 (Rayleigh Problem)

Minimize the cost functional

$$J = \int_0^{2.5} (x_1^2 + u^2)\, dt$$

for the second-order system

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = -x_1 + 1.4\, x_2 - 0.14\, x_2^3 + 4u$$

subject to the initial and terminal conditions

$$x_1(0) = -5,\ x_2(0) = -5,\ x_1(2.5) = 1.5,\ x_2(2.5) = 2.$$

For this problem approximate solutions were obtained for several values of N. The minimum cost attained in each case is shown in Table 4.1 below.

The cubic spline parametrization (4.24) means that for each N, q is an (N+3)-dimensional parameter vector. There are two initial and two terminal conditions given here, so the problem can be solved as a static optimization problem in N+3 variables subject to four linear constraints. Alternatively, these four constraints could be used to eliminate four components of the parameter vector q, in which case we have an unconstrained static optimization in N-1 variables.

For our present computations, the initial conditions were used to eliminate two components of q. So we have a static optimization problem in N+1 variables subject to two linear constraints. Feasible values for the remaining components could be obtained by minimizing the error function $\langle \tilde{\eta}, \tilde{\eta} \rangle$ using the Davidon-Fletcher-Powell algorithm. But for this problem they can also be obtained by simple hand calculations. The Goldfarb-Lapidus ensures that the two terminal conditions were being satisfied at at all times; in our computations we have

$$|\tilde{\eta}_i(q)| < 10^{-9} \quad \text{for } i = 1,2.$$

For the sake of comparison, the minimum cost obtained by Miele et. al. [6] for an equivalent problem is 29.377. This problem has also been solved by Lastman [8] using a Chebyshev polynomial parametrization of the control.

TABLE 4.1 MINIMUM COST FOR EACH N

N	$J(\bar{x}_1)$
2	33.7502
4	29.4502
6	29.4619
8	29.4026
10	29.3935
12	29.3930
15	29.3845
20	29.3814

The cost J was evaluated numerically by applying the 10-point Gauss-Legendre quadrature rule to each section in the mesh, and is therefore accurate.

Example 2 (Van der Pol Problem)

Minimize the cost functional

$$J = \frac{1}{2} \int_0^5 (x_1^2 + x_2^2 + u^2)\, dt$$

for the second-order system

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = -x_1 + (1 - x_1^2)\, x_2 + u$$

subject to the linear constraints

$$x_1(0) = 1, \quad x_2(0) = 0, \quad x_1(5) - x_2(5) + 1 = 0.$$

The computational details are similar to those of Example 1. Approximate solutions were obtained for several values of N, and the minimum cost attained in each case is shown in Table 4.2 below. The control and state profiles obtained for N = 10 are summarised in Table 4.3.

TABLE 4.2 MINIMUM COST FOR EACH N

N	$J(\bar{x}_1)$
2	1.90834
4	1.69895
6	1.68821
8	1.68643
10	1.68597
15	1.68574
20	1.68570

TABLE 4.3 CONTROL AND STATE PROFILES (N = 10)

t	$\bar{x}_1(t)$	$\bar{x}_2(t)$	$\bar{u}(t)$
0.0	0.1000E1	0.0000	-0.3428E0
0.5	0.8719E0	-0.4327E0	0.5877E0
1.0	0.6245E0	-0.5221E0	0.9734E0
1.5	0.3723E0	-0.4765E0	0.9349E0
2.0	0.1524E0	-0.4046E0	0.6829E0
2.5	-0.3555E-1	-0.3525E0	0.3898E0
3.0	-0.2035E0	-0.3212E0	0.1563E0
3.5	-0.3535E0	-0.2703E0	0.3462E-1
4.0	-0.4587E0	-0.1285E0	0.5863E-1
4.5	-0.4536E0	0.1836E0	0.2330E0
5.0	-0.2292E0	0.7708E0	0.5569E0

4.5 CONCLUSIONS

The SP procedure for solving the general optimal control problem has been described. For multivariable systems that can be cast into the phase variable form (4.21), and this includes the class of all controllable linear multivariable systems, it has been shown that the basic SP procedure could be suitably modified to improve its efficiency. The implementation of this modified SP procedure on problems involving linear terminal constraints employing spline approximation spaces was discussed, and approximate solutions to two sample nonlinear problems were obtained using cubic splines.

REFERENCES

[1] R. K. Mehra and R. E. Davis: A Generalized Gradient Method for Optimal Control Problems with Inequality Constraints and Singular Arcs, IEEE Trans. Automatic Control, AC-17 (1972) pp 69-79.

[2] C. D. Johnson and W. M. Wonham: A Note on the Transformation to Canonical (Phase-Variable) Form, IEEE Trans. Automatic Control, AC-9 (1964) pp 312-313.

[3] L. M. Silverman: Transformation of Time Variable Systems to Canonical (Phase-Variable) Form, IEEE Trans. Automatic Control, AC-11 (1966) pp 300-303.

[4] D. G. Luenberger: Canonical Forms for Linear Multivariable Systems, IEEE Trans. Automatic Control, AC-12 (1967) pp 290-293.

[5] D. Goldfarb and L. Lapidus: Conjugate Gradient Method for Nonlinear Programming Problems with Linear Constraints, Ind. Engng. Chem. Fundamentals, Vol.7 (1968) pp 142-151.

[6] A. Miele, J. L. Tietze and A. V. Levy: Summary and Comparison of Gradient Restoration Algorithms for Optimal Control Problems, J. Optimiz. Theory Appl., Vol.10 (1972) pp 381-403.

[7] R. J. O'Doherty and B. L. Pierson: A Numerical Study of Augmented Penalty Function Algorithms for Terminally Constrained Optimal Control Problems, J. Optimiz. Theory Appl., Vol.14 (1974) pp 393-403.

[8] G. J. Lastman: Sub-optimal Open Loop Control of Nonlinear Systems Using Approximations for the Controls, Int. J. Control, Vol.20 (1974) pp 289-303.

[9] J. B. Rosen: The Gradient Projection Method for Nonlinear Programming II: Nonlinear Constraints, SIAM J. Appl. Math., Vol. 9 (1961) pp 514-532.

CHAPTER 5

ERROR BOUNDS FOR THE STATE PARAMETRIZATION PROCEDURE

5.1 INTRODUCTION

The aim of the present chapter is to establish error bounds for the SP approximation over spline spaces. The class of problems to be considered here shall be that described in section 4.3. Firstly, we shall restrict our attention to the linear quadratic problem.

One possible way of obtaining error bounds is to proceed along a similar line as that taken in section 3.2 in obtaining error estimates for the CP approximation. In this approach, the key to the derivation of error estimates is the fundamental identity (3.7) concerning the cost functional. However, adopting this approach for deriving error estimates of the SP approximation for the class of problems being considered will lead to sub-optimal error bounds for the state variables.

For instance, let us consider the double integrator system. We note that in this case, the SP approximation over the cubic spline space is identical to the CP approximation over the linear spline space. It has been observed from the numerical results presented in section 3.2 that the error bounds in the L_2-norm for the approximations $\bar{x}_1^h$ and $\bar{x}_2^h$ should be $O(h^4)$ and $O(h^3)$ respectively, whereas Theorem 3.2 could predict only $O(h^2)$ convergence rate for both $\bar{x}_1^h$ and $\bar{x}_2^h$. This suggests that the error bounds contained in Theorem 3.2 are probably not optimal for simple pure integrator systems.

Fortunately, by adopting a different approach for deriving the error bounds, the expected higher order bounds for the state variables can be established not only for pure integrator systems, but more generally for linear systems of the phase variable canonical form. In the derivation that will be presented in section 5.3 we employ known results concerning the Ritz procedure solution of variational problems over spline approximation spaces. We shall now recall the relevant details concerning the Ritz procedure.

5.2 THE RITZ PROCEDURE

Let H_o denote the subspace of the Sobolev space $\{W_2^n[0,T],E^1\}$ whose members satisfy the homogeneous boundary conditions

$$D^i w(0) = D^i w(T) = 0 , \qquad i = 0,1,\ldots, n-1 \tag{5.1}$$

where the symbol $D^i w$ denotes the i-fold differentiation of w with respect to t.

Let π be a continuous bilinear form on H_o given by,

$$\pi(w_1,w_2) \equiv \int_0^T \sum_{i=1}^{n} \sum_{j=1}^{n} \mu_{ij}(t)\, D^i w_1\, D^j w_2 \tag{5.2}$$

for all $w_1, w_2 \in H_o$.

Suppose that π is an elliptic bilinear form; that is, we assume there exists a constant $\sigma > 0$ such that

$$\sigma ||w||_{2,n}^2 \leq \pi(w,w) \tag{5.3}$$

for all $w \in H_o$.

Let L be a continuous linear form on H_o given by

$$L(w) \equiv \int_0^T \sum_{i=1}^{n} \nu_i(t)\, D^i w \, dt. \tag{5.4}$$

Consider now the minimization of the quadratic functional

$$J(w) \equiv \pi(w,w) - L(w) \tag{5.5}$$

over H_o. We seek the element $w^* \varepsilon H_o$ such that

$$J(w^*) = \inf\{J(w) \,|\, w \varepsilon H_o\}. \tag{5.6}$$

The existence of a unique solution w^* to the above variational problem is ensured by the ellipticity assumption on π (see [2]).

The Ritz procedure for solving the variational problem consists of choosing a finite-dimensional subspace $C_m \subset H_o$ and determining the element $\bar{w}_m \varepsilon C_m$ such that

$$J(\bar{w}_m) = \inf\{J(w_m) \,|\, w_m \varepsilon C_m\}. \tag{5.7}$$

The existence of a unique solution $\bar{w}_m$ for every $C_m \subset H_o$ is also ensured by the ellipticity assumption on π (see [2]).

For our present purpose we shall only be interested in the case where C_m is a spline space S_h^α with a uniform mesh. A result which we will need later is the following theorem (see [1], [2]) which specifies error bounds for the Ritz approximation $\bar{w}^h$ and its derivatives.

Theorem 5.1

Suppose that the bilinear form π defined by (5.2) is elliptic, and suppose that the functions μ_{ij} are arbitrarily smooth. Let w^* denote the solution to the variational problem (5.5), (5.6), and let $\bar{w}^h$ denote the corresponding Ritz solution over S_h^α, in which the partition is assumed to be uniform. Then the accuracy of $\bar{w}^h$ is given by the following error bounds:

$$\left|\left|\bar{w}^{h} - w^{*}\right|\right|_{2,p} = \begin{cases} Ch^{\alpha-p} \left|\left|w^{*}\right|\right|_{2,\alpha} & \text{if } p \geq 2n - \alpha \\ Ch^{2(\alpha-n)} \left|\left|w^{*}\right|\right|_{2,\alpha} & \text{if } p \leq 2n - \alpha \end{cases} \tag{5.8}$$

The exponents in (5.8) are optimal, so the order of accuracy never exceeds $2(\alpha-n)$ in any norm.

Remarks:

The proof of the above theorem is given in Strang and Fix [1] and Schultz [2]. Its derivation involves the use of an ingenious mathematical argument known in numerical analysis literature as Nitsche's trick. The applicability of Nitsche's trick depends on π being strongly coercive (see [2] for definition) over H_o. This condition is satisfied when we require μ_{ij} to be arbitrarily smooth.

5.3 LINEAR QUADRATIC PROBLEMS

We can now proceed with the derivation of error estimates for the SP approximation over spline spaces. For the moment we shall confine our attention to linear quadratic problems involving scalar controls. Extensions to linear multivariable problems and nonlinear problems will be considered later.

Problem Statement

Consider the optimal control problem described by the linear system expressed in phase variable form

$$\begin{aligned} \dot{x}_1 &= x_2 \\ &\vdots \\ \dot{x}_{n-1} &= x_n \\ \dot{x}_n &= a_1(t)\, x_1 + \ldots + a_n(t)\, x_n + u \end{aligned} \tag{5.9}$$

and the quadratic cost functional

$$J = \int_0^T \left[\langle x, Q(t)x \rangle + R(t)\, u^2 \right] dt \tag{5.10}$$

where $Q(t)$ is an $n \times n$ symmetric, positive semi-definite matrix and $R(t)$ is a strictly positive scalar-valued function. It is assumed that $a_1, \ldots, a_n$, Q and R are arbitrarily smooth functions. The final time T is assumed to be fixed, and the state vector is assumed to be specified at the initial and final times.

Remarks

The above assumptions imply that the optimal control u^* and state x^* are arbitrarily smooth. We could have assumed that the above problem was finitely smooth, as was done in Chapter 3. The analysis and results will remain unchanged except for some minor and obvious modifications.

Change of Variables

We have assumed that the state vector is fixed at the initial and final times. In general these boundary conditions will be non-zero but it is not hard to see that provided we have a sufficiently fine mesh we can find a spline $s(t) \in S_h^\alpha$, where $\alpha > n$, such that the new state variables defined by

$$z_1 \equiv x_1 - s(t), \quad z_2 \equiv \dot{z}_1, \ldots, z_n \equiv \dot{z}_{n-1} \tag{5.11}$$

will satisfy the homogeneous boundary conditions

$$z(0) = z(T) = 0 , \tag{5.12}$$

where $z \equiv \left[z_1, z_2, \ldots, z_n \right]^T$.

Let us define a new control variable v by

$$v(t) \equiv u(t) + \sum_{i=1}^{n} a_i D^{i-1} s(t) - D^n s(t). \tag{5.13}$$

When expressed in terms of the new state variables $z_1,\ldots,z_n$ and the new control variable v, the system equation (5.9) takes on the form

$$\begin{aligned} \dot{z}_1 &= z_2 \\ &\vdots \\ \dot{z}_{n-1} &= z_n \\ \dot{z}_n &= a_1 z_1 + \ldots + a_n z_n + v \end{aligned} \tag{5.14}$$

where the state vector satisfies the homogeneous boundary conditions given by (5.12).

Expressing the cost functional J of (5.10) in terms of the new variables z and v, it is easily seen that the new expression for J will consist of a linear portion as well as a quadratic portion, plus a constant term that can be omitted. It is also clear that the quadratic portion retains the same form as before. Thus,

$$J = \int_0^T \left[\langle z, Qz \rangle + Rv^2 + \langle c, z \rangle + gv \right] dt \tag{5.15}$$

where c and g are known functions of Q, R and s.

Suppose now that the SP procedure is applied to the new problem given by (5.14) and (5.15). The state variable z_1 is taken as the independent variable and parametrized as an element of S_h^α over a uniform partition of $[0,T]$. The SP solution $\bar{z}_1^h$ is given by

$$J(\bar{z}_1^h) = \inf\{J(z_1^h) \mid z_1^h \in S_h^\alpha\} \tag{5.16}$$

subject to the initial and terminal conditions of (5.12).

Let $\bar{x}_1^h$ be the SP solution over S_h^α to the problem given by (5.9) and (5.10). It is easy to see that $\bar{x}_1^h$ and $\bar{z}_1^h$ are related by the equation

$$\bar{x}_1^h = \bar{z}_1^h + s(t). \tag{5.17}$$

THE ELLIPTICITY CONDITION

We shall now apply Theorem 5.1 to obtain order bounds for the error $\bar{z}_1^h - z_1^*$ and its derivatives. In order to apply this theorem we must consider J as a functional in the variable z_1; i.e. we view $z_2,\ldots,z_n$ as derivatives of z_1, and view the control v as a function of z_1 and its derivatives. We write the quadratic portion of J as

$$\pi(z_1,z_1) \equiv \int_0^T \left[\langle z, Qz \rangle + Rv^2 \right] dt. \tag{5.18}$$

We must now show that the above quadratic functional is elliptic : i.e. there exists a constant $\sigma > 0$ such that

$$\pi(z_1,z_1) \geq \sigma ||z_1||_{2,n}^2 \quad \text{for all } z_1 \in H_o. \tag{5.19}$$

The demonstration of this ellipticity condition is quite straightforward. First of all, we can drop the positive semi-definite part of $\pi(z_1,z_1)$ and write

$$\pi(z_1,z_1) \geq \int_0^T R(t)\, v^2\, dt. \tag{5.20}$$

Since R(t) has been assumed to be a smooth, positive function over $[0,T]$, the minimum of R(t) exists and is clearly positive: i.e.

$$R_o \equiv \min\{R(t) \mid t \in [0,T]\} > 0. \tag{5.21}$$

Hence, it follows from (5.20) that

$$\pi(z_1,z_1) \geq R_o ||v||_2^2. \tag{5.22}$$

Now, application of the Gronwall Inequality to the system equation (5.14) yields the result that

$$||z_i||_2 \leq c_1 ||v||_2 , \qquad i = 1,\ldots,n \tag{5.23}$$

for some constant $c_1 > 0$. From (5.14) it also follows that for some $c_2 > 0$,

$$||\dot{z}_n||_2^2 \leq c_2\left[\, ||z_1||_2^2 + \ldots + ||z_n||_2^2 + ||v||_2^2 \,\right] \tag{5.24}$$

which, using (5.23), can be reduced to

$$||\dot{z}_n||_2^2 \leq c_3||v||_2^2 \tag{5.25}$$

for some $c_3 > 0$.

Finally, it is easy to deduce from (5.23) and (5.25) that

$$||z_1||_{2,n}^2 \equiv ||z_1||_2^2 + \ldots + ||z_n||_2^2 + ||\dot{z}_n||_2^2 \leq c||v||_2^2 \tag{5.26}$$

for some $c > 0$, which may then be combined with (5.22) to yield the required ellipticity condition.

ERROR ESTIMATES

We can now apply Theorem 5.1 to the inner-product π defined by (5.18) to obtain

$$||\bar{z}_1^h - z_1^*||_{2,p} = O(h^{\alpha-p} + h^{2(\alpha-n)}). \tag{5.27}$$

In view of (5.17) we also obtain

$$||\bar{x}_1^h - x_1^*||_{2,p} = O(h^{\alpha-p} + h^{2(\alpha-n)}). \tag{5.28}$$

We note that the error bound (5.28) tells us that for convergence to occur we must have $\alpha > n$. That is, if our control problem involves an n-dimensional system, then the state variable x_1 must be parametrized as a spline of order n or higher.

From the above error bounds for $\bar{x}_1^h$ and its derivatives, the corresponding error bounds for the control $\bar{u}^h$ and the cost $J(\bar{x}_1^h)$ can be readily deduced to be of orders $O(h^{\alpha-n})$ and $O(h^{2(\alpha-n)})$ respectively.

We summarise the results in the following theorem.

Theorem 5.2

Let $(\bar{u}^h, \bar{x}^h)$ be the control-state pair prescribed by the SP procedure for the problem given by (5.9), (5.10) and the associated assumptions, over the spline approximation space S_h^α with a uniform partition. Then

$$||\bar{u}^h - u^*||_2 \leq O(h^{\alpha-n}) \tag{5.29}$$

$$||\bar{x}_i^h - x_i^*||_2 \leq O(h^{\alpha+1-i} + h^{2(\alpha-n)}), \quad i = 1,\ldots,n \tag{5.30}$$

and

$$0 \leq J(\bar{x}_1^h) - J(x_1^*) \leq O(h^{2(\alpha-n)}). \tag{5.31}$$

Remarks

Although we assumed in the above theorem that the initial and final values of the state vector are completely specified, the error bounds in the theorem remain valid in the case when the initial value of the state is specified together with a general linear terminal constraint. Needless to say, this includes the special case where the terminal constraint is absent.

Numerical Example

Approximate solutions to the following linear quadratic problem were obtained using the SP method.

For the second order system

$$\dot{x}_1 = x_2 \qquad x_1(o) = 0$$

$$\dot{x}_2 = -x_2 + u \qquad x_2(0) = -1$$

find the control u* that minimizes the cost functional

$$J = \int_0^1 (x_1^2 + x_2^2 + .005u^2)\, dt.$$

The analytical solution of the above optimal control problem was found to be given by

$$x_1^* = a_1 e^{\beta t} + a_2 e^{-\beta t} + a_3 e^{t} + a_4 e^{-t} \tag{5.32}$$

where $\beta = 10\sqrt{2}$, and a_1, a_2, a_3, a_4 are constants given by

$$a_1 = 0.162,208 \times 10^{-9} , \qquad a_2 = 0.074,735 ,$$

$$a_3 = -0.008,914 , \qquad a_4 = -0.065,821.$$

The optimal variables x_2^* and u* can be easily obtained from the state equations. The optimal cost J* was computed to be

$$J^* = 0.069,361.$$

The state variable x_1 was parametrized as a cubic spline; this ensures that the resulting control will be continuous. A uniform partition of $[0,1]$ was used and results were obtained for several values of N, the number of sections in the partition. For each N, the mesh size is h = 1/N. The definite integrals for this example were evaluated by applying the 4-point Gauss-Legendre quadrature formula over each section in the partition.

TABLE 5.1 CONVERGENCE HISTORIES

N	$J(\bar{x}_1^h)$	$J(\bar{x}_1^h)-J^*$	$\|\|\bar{x}_1^h-x_1^*\|\|_2$	$\|\|\bar{x}_2^h-x_2^*\|\|_2$	$\|\|\bar{u}^h-u^*\|\|_2$
2	.098 437	.029 076	.249E-1	.123E0	.163E1
3	.081 302	.011 941	.104E-1	.686E-1	.119E1
4	.074 781	.005 420	.466E-2	.393E-1	.878E0
5	.072 075	.002 714	.233E-2	.238E-1	.655E0
6	.070 833	.001 472	.126E-2	.151E-1	.499E0
8	.069 883	.000 522	.440E-3	.690E-2	.308E0
10	.069 583	.000 223	.190E-3	.360E-2	.205E0

The above results were used to construct the following table of convergence rates, where α_N, β_N, γ_N, and δ_N are as defined in section 3.2.

TABLE 5.2 CONVERGENCE RATES

N	α_N	β_N	γ_N	δ_N
3	2.20	2.16	1.45	0.76
4	2.74	2.77	1.93	1.07
5	3.11	3.11	2.27	1.31
6	3.34	3.37	2.51	1.48
8	3.60	3.65	2.73	1.67
10	3.81	3.88	2.91	1.82

From Theorem 5.2, the following convergence rates are predicted for the cubic spline approximation space S_h^4:

$$0 \leq J(\bar{x}_1^h) - J^* \leq O(h^4)$$

$$||\bar{x}_1^h - x_1^*||_2 \leq O(h^4 + h^4) = O(h^4)$$

$$||\bar{x}_2^h - x_2^*||_2 \leq O(h^3 + h^4) = O(h^3)$$

and

$$||\bar{u}^h - u^*||_2 \leq O(h^2).$$

We observe that these convergence rates agree well with the figures in Table 5.2 as N increases. The optimal and approximate (for N = 5,10) solution profiles are summarised in Tables 5.3 to 5.5 below.

TABLE 5.3 CONTROL PROFILES

t	N = 5 $\bar{u}(t)$	N = 10 $\bar{u}(t)$	$u^*(t)$
0.0	.9495E1	.1219E2	.1387E2
0.1	.4894E1	.2637E1	.3357E1
0.2	-.2456E0	.7193E0	.7992E0
0.3	-.2085E-1	.1290E0	.1755E0
0.4	.2279E0	.1869E-1	.2194E-1
0.5	.6680E-1	-.2077E-1	-.1755E-1
0.6	-.1116E0	-.2950E-1	-.2945E-1
0.7	-.6394E-1	-.3479E-1	-.3451E-1
0.8	-.1045E-1	-.3685E-1	-.3664E-1
0.9	-.2333E-1	-.3437E-1	-.3203E-1
1.0	-.3725E-1	-.5314E-2	.0000E0

TABLE 5.4 STATE TRAJECTORIES (x_2)

t	N = 5 $\bar{x}_2(t)$	N = 10 $\bar{x}_2(t)$	$x_2^*(t)$
0.0	-.1000E1	-.1000E1	-.1000E1
0.1	-.2196E0	-.1989E0	-.2073E0
0.2	.2271E-1	-.2010E-1	-.1947E-1
0.3	.7864E-2	.2221E-1	.2154E-1
0.4	.1697E-1	.2713E-1	.2713E-1
0.5	.2939E-1	.2445E-1	.2433E-1
0.6	.2445E-1	.1972E-1	.1967E-1
0.7	.1377E-1	.1478E-1	.1473E-1
0.8	.8912E-2	.9964E-2	.9914E-2
0.9	.6455E-2	.5624E-2	.5611E-2
1.0	.2955E-2	.3199E-2	.3183E-2

TABLE 5.5 STATE TRAJECTORIES (x_1)

t	N = 5 $\bar{x}_1(t)$	N = 10 $\bar{x}_1(t)$	$x_1^*(t)$
0.0	.0000E0	.0000E0	.0000E0
0.1	-.5649E-1	-.5132E-1	-.5124E-1
0.2	-.6185E-1	-.6052E-1	-.6036E-1
0.3	-.6052E-1	-.5989E-1	-.5972E-1
0.4	-.5948E-1	-.5732E-1	-.5716E-1
0.5	-.5702E-1	-.5471E-1	-.5456E-1
0.6	-.5418E-1	-.5250E-1	-.5235E-1
0.7	-.5232E-1	-.5078E-1	-.5063E-1
0.8	-.5123E-1	-.4954E-1	-.4940E-1
0.9	-.5045E-1	-.4877E-1	-.4863E-1
1.0	-.4998E-1	-.4835E-1	-.4822E-1

Remarks

We note that the error bounds prescribed by Theorem 5.2 are also consistent with the numerical results presented in Chapter 3 for the CP solution. The numerical examples used involve pure integrator systems, so in these cases the CP and SP solutions are identical (assuming appropriate approximation spaces are employed).

5.4 MULTIVARIABLE SYSTEM PROBLEMS

We now consider the extension of the previous results to include problems involving linear multivariable systems.

Problem Statement

We consider here the class of multivariable systems composed of a coupled set of single-input subsystems each of which is of the form

$$\begin{aligned}
\dot{x}_{m_i+1} &= x_{m_i+2} \\
\dot{x}_{m_i+2} &= x_{m_i+3} \\
&\vdots \\
\dot{x}_{m_i+n_i} &= \langle a_i(t), x\rangle + u_i \,, \qquad i = 1,2,\ldots,r
\end{aligned} \tag{5.33}$$

where $m_1 \equiv 0$ and $m_i \equiv n_1+\ldots+n_{i-1}$ for $i \geq 2$. Let $n \equiv n_1+\ldots+n_r$; x is the n-dimensional state vector given by $x \equiv [x_1,\ldots,x_n]^T$, and each $a_i(t)$ is an n-dimensional vector.

Let the cost functional be defined by

$$J = \int_0^T [\langle x, Q(t)x\rangle + \langle u, R(t)u\rangle]\,dt \tag{5.34}$$

where Q is an $n \times n$ symmetric, positive semi-definite matrix, R is an $r \times r$ symmetric, positive definite matrix and $u \equiv [u_1,\ldots,u_r]^T$ is the r-dimensional control vector.

It is assumed that $a_1,\ldots,a_r$, Q and R are arbitrarily smooth. We also assume that the initial condition for x is specified and that terminal constraints may or may not be present.

The Parametrization

As described in Chapter 4, we choose $x_{m_1+1},\ldots,x_{m_r+1}$ as independent variables and each of them is individually parametrized. Using spline approximation spaces this means that we require $x_{m_1+1} \in S_{h_1}^{\alpha_1},\ldots,x_{m_r+1} \in S_{h_r}^{\alpha_r}$, where $\alpha_1 > n_1,\ldots,\alpha_r > n_r$, subject to given boundary conditions. This parametrization is reasonable for computational purposes, but for the sake of obtaining error bounds, the following parametrization procedure is assumed to be adopted.

We have r subsystems of dimensions $n_1,\ldots,n_r$; suppose that

$$n_o = \max\{n_1,\ldots,n_r\}. \tag{5.35}$$

We choose an $\alpha > n_o$, and adopt the following parametrization of the independent variables:

$$x_{m_i+1} \in S_h^{\alpha-(n_o-n_i)}, \qquad i = 1,\ldots,r \tag{5.36}$$

subject to given boundary conditions, where the same uniform mesh applies to all the independent variables.

This is equivalent to introducing higher order auxiliary variables within each subsystem so that each subsystem is extended to become n_o-dimensional, and then parametrizing the highest order variable of each subsystem as an element of S_h^{α}. To illustrate, suppose that the first of the subsystems in (5.33) is of dimension $n_1 < n_o$. By introducing additional variables $z_1,\ldots,z_{n_o-n_1}$, this subsystem can be written

$$
\begin{aligned}
\dot{z}_1 &= z_2 \\
\dot{z}_2 &= z_3 \\
&\vdots \\
\dot{z}_{n_o-n_1} &= x_1 \\
\dot{x}_1 &= x_2 \\
&\vdots \\
\dot{x}_{n_1} &= \langle a_1(t), x\rangle + u_1
\end{aligned}
\tag{5.37}
$$

in which the initial conditions for the x variables are given and the initial conditions for the z variables can be arbitrarily specified. The variable z_1 is taken to be the independent variable and parametrized by requiring $z_1 \in S_h^\alpha$, where $\alpha > n_o$.

This procedure can be carried out for all those subsystems whose dimensions are less than n_o.

ERROR ESTIMATES

We note that the result contained in Theorem 5.1 applies also to the vector case where H_o is a subspace of $\{W_2^n[0,T], E^r\}$, with $r > 1$ (see [1]). Thus the analysis of the preceding section may be extenced to the multivariable case; we summarise the result in the following theorem.

Theorem 5.3

Consider the problem defined by (5.33) and (5.34). Let $(\bar{u}^h, \bar{x}^h)$ be the control-state pair prescribed by the SP procedure for the problem, where the parametrization of (5.36) is assumed. Then for each

$i = 1,\ldots,r,$

$$\left|\left|\bar{u}_i^h - u^*\right|\right|_2 \leq O(h^{\alpha-n_o}) \tag{5.38}$$

$$\left|\left|\bar{x}^h_{m_i+j} - x^*_{m_i+j}\right|\right|_2 \leq O(h^{\alpha+1-j-(n_o-n_i)} + h^{2(\alpha-n_o)})$$

$$\text{for } j = 1,\ldots,n_i \tag{5.39}$$

and

$$0 \leq \bar{J}^h - J^* \leq O(h^{2(\alpha-n_o)}) \ , \tag{5.40}$$

where $\bar{J}^h \equiv J(\bar{x}^h_1, \bar{x}^h_{m_2+1},\ldots,\bar{x}^h_{m_r+1})$.

5.5 NONLINEAR PROBLEMS

We now consider the convergence of the SP procedure for the nonlinear optimal control problem. In the vicinity of the optimum the nonlinear system behaves in an almost linear fashion while the cost functional is nearly quadratic in shape. Therefore it is reasonable to presume that the error bounds of Theorem 5.2 will eventually be valid as the SP approximations get sufficiently close to the optimum; in other words, when the mesh parameter tends to zero the error bounds of Theorem 5.2 will hold.

Problem Statement

Consider now the problem described by the state equation

$$\begin{aligned} \dot{x}_1 &= x_2 \\ &\vdots \\ \dot{x}_{n-1} &= x_n \\ \dot{x}_n &= f(x,u,t) \ , \qquad x(0) \text{ given,} \end{aligned} \tag{5.41}$$

and the cost functional

$$J = \int_0^T \phi(x,u,t)\, dt \ . \tag{5.42}$$

Let $H(x,u,\lambda,t)$ be the Hamiltonian defined in the usual way,

$$H(x,u,\lambda,t) \equiv \phi(x,u,t) + \lambda_1 x_2 + \ldots + \lambda_{n-1} x_n + \lambda_n f(x,u,t). \tag{5.43}$$

The following assumptions are made on the problem:

(A1) $f(x,u,t)$ and $\phi(x,u,t)$ are infinitely differentiable in x, u and t.

(A2) The second variation of the cost functional J is strongly positive at the optimum; i.e., for some $\sigma > 0$,

$$\int_0^T \langle H''(x^*,u^*,\lambda^*,t) \begin{bmatrix}\delta u \\ \delta x\end{bmatrix}, \begin{bmatrix}\delta u \\ \delta x\end{bmatrix} \rangle\, dt \geq \sigma \|\delta u\|_2^2 \tag{5.44}$$

for all δx and δu satisfying

$$\begin{aligned} \delta\dot{x}_1 &= \delta x_2 \\ &\vdots \\ \delta\dot{x}_{n-1} &= \delta x_n \\ \delta\dot{x}_n &= \langle f_x^*, \delta x\rangle + f_u^* \delta u \ , \qquad \delta x(0) = 0 \ , \end{aligned} \tag{5.45}$$

where

$$H''(x^*,u^*,\lambda^*,t) \equiv \begin{bmatrix} H^*_{uu} & H^*_{ux} \\ H^*_{xu} & H^*_{xx} \end{bmatrix} ,$$

and we have adopted the notations

$$f_u^* \equiv \frac{\partial f}{\partial u}(x^*,u^*,t), \quad H^*_{uu} \equiv \frac{\partial^2 H}{\partial u^2}(x^*,u^*,\lambda^*,t) \ , \text{ etc.}$$

(A3) $f_u^* \neq 0$ for all $t \in [0,T]$.

Linear Quadratic Approximation

We now assume that in the vicinity of the optimum the original problem of (5.41) and (5.42) is adequately described by the linearized version of (5.41) and the quadratic approximation to (5.42). The linearized system is

$$\begin{aligned} \dot{x}_1 &= x_2 \\ &\vdots \\ \dot{x}_{n-1} &= x_n \\ \dot{x}_n &= f^* + \langle f^*_x, x-x^* \rangle + f^*_u (u-u^*) , \end{aligned} \tag{5.46}$$

and by expanding the performance index to second order the approximation for the cost is

$$J = J^* + \int_0^T \langle H''(x^*,u^*,\lambda^*,t) \begin{bmatrix} u-u^* \\ x-x^* \end{bmatrix}, \begin{bmatrix} u-u^* \\ x-x^* \end{bmatrix} \rangle \, dt. \tag{5.47}$$

So the linear quadratic problem that we have now is to minimize J given by (5.47) subject to the system equation (5.46).

As in section 5.3 we can define a new state vector z and a new control v so that the system equation (5.46) takes the form

$$\begin{aligned} \dot{z}_1 &= z_2 \\ &\vdots \\ \dot{z}_{n-1} &= z_n \\ \dot{z}_n &= \langle f^*_x, x \rangle + f^*_u . v , \qquad z(0) = 0 , \end{aligned} \tag{5.48}$$

and the cost functional to be minimized becomes

$$\begin{aligned} J_2 \equiv J-J^* &= \int_0^T \langle H''(x^*,u^*,\lambda^*,t) \begin{bmatrix} v-v^* \\ z-z^* \end{bmatrix} , \begin{bmatrix} v-v^* \\ z-z^* \end{bmatrix} \rangle \, dt. \\ &= \int_0^T \langle H''(x^*,u^*,\lambda^*,t) \begin{bmatrix} v \\ z \end{bmatrix} , \begin{bmatrix} v \\ z \end{bmatrix} \rangle \, dt \\ &\quad + \text{linear portion} + \text{constant}. \end{aligned} \tag{5.49}$$

Again, as in section 5.3, we denote the quadratic portion of J_2 by $\pi(z_1,z_1)$:

$$\pi(z_1,z_1) \equiv \int_0^T \langle H''(x^*,u^*,\lambda^*,t) \begin{bmatrix} v \\ z \end{bmatrix}, \begin{bmatrix} v \\ z \end{bmatrix} \rangle \, dt. \tag{5.50}$$

This is of the same form as (5.18) except that now we have a cross product vz term as well. However, what is important is that we can use assumption (A2) to say that

$$\pi(z_1,z_1) \geq \sigma ||v||_2^2 \quad \text{for some } \sigma > 0, \tag{5.51}$$

and the ellipticity of π can then be deduced using the arguments of section 5.3.

It follows that we can expect the results of Theorem 5.2 to be valid (asymptotically) for the nonlinear case.

Remarks

Let us take a look at the numerical results of Examples 1 and 2 given in Chapter 4. These are nonlinear problems for which the analytical solutions are not known, so we cannot plot tables of convergence histories or convergence rates. However, an examination of the cost figures for the Van der Pol problem reveals that they are consistent with an $O(h^4)$ convergence rate with the optimum cost about 1.68568. This order of convergence is what we would expect from the theory. For the Rayleigh problem, the cost figures exhibit a more erratic behaviour. A possible explanation for this behaviour is that the asymptotic range for this problem has not been reached.

5.6 CONCLUSIONS

Convergence of the SP procedure over spline approximation spaces has been investigated for a wide class of optimal control problems. Sharp error bounds for the control, state and cost functional have been derived by using known results regarding the Ritz solution of variational problems involving elliptic quadratic functionals over spline approximation spaces.

The error bounds have been initially obtained in the mean square norm for the linear quadratic problem with a scalar control. These results were then extended to cover more general cases, viz. problems with multivariable controls and problems which are nonlinear. The validity of these error bounds have been seen to be supported by the numerical evidence presented.

REFERENCES

[1] G. Strang and G. J. Fix: An Analysis of the Finite Element Method, Prentice-Hall, Inc., Englewood Cliffs, N.J. (1973).

[2] M. H. Schultz: L^2 Error Bounds for the Rayleigh-Ritz-Galerkin Method, SIAM J. Numer. Anal., Vol.8 (1971) pp.737-748.

CHAPTER 6

THE STATE PARAMETRIZATION PROCEDURE FOR DISTRIBUTED PARAMETER SYSTEMS

6.1 INTRODUCTION

In previous chapters we studied the CP and the SP procedures for solving optimal control problems involving lumped systems. Both procedures can be generalized in a straightforward manner to handle control problems involving distributed controls. However, the CP procedure for distributed system problems would involve the solution of partial differential equations for the state and co-state. In comparison to the numerical solution of ordinary differential equations, the numerical solution of nonlinear partial differential equations represents a formidable task with considerable computational requirements. Hence in general the CP procedure will not provide a satisfactory means of solving distributed control problems. An exception to this rule occurs when we are dealing with a linear distributed system for which the Green's function can be constructed. In Appendix C the CP procedure is used to solve a linear distributed system problem involving a boundary control. Nevertheless we shall not be considering the CP procedure here.

In this chapter we shall extend the SP procedure to problems involving systems of the form

$$\frac{\partial x}{\partial t} = f(x, \frac{\partial x}{\partial y}, \ldots, \frac{\partial^k x}{\partial y^k}, u) \tag{6.1}$$

in which the state $x(y,t)$ and control $u(y,t)$ are scalar variables.

We remark here that the SP procedure is applicable to a wider class of problems than is indicated by (6.1). However, for the purpose of illustrating the computational procedure we have chosen the system equation (6.1). Later on in section 6.4 we apply the procedure to a problem involving a nonlinear system of the hyperbolic type which does not fit into the form (6.1).

In the following section 6.2 we discuss the classification of control systems into types, and indicate the basic difference between parabolic and hyperbolic systems. Although the applicability of the SP procedure does not depend on the type of the control system, this section has been included for the sake of completeness. On the other hand we have not gone into any depth in our rather sketchy discussion; detailed treatments of this subject can be found in [1], [2].

In section 6.3 we summarise the essential approximation properties of bivariate splines, and following that, section 6.4 contains a description of the SP procedure for a general class of problems, as well as computation results for two specific examples. The question of convergence of the SP solution employing multivariate spline approximation spaces as the mesh size decreases is examined in section 6.5 for a class of second order linear system problems with quadratic performance indices under appropriate assumptions.

6.2 CLASSIFICATION OF DISTRIBUTED SYSTEMS

In the general formulation of the lumped optimal control problem the state equation is written as a set of first order ordinary differential equations. However, it is usual to find in applications that the system description actually comes in the form of one or more

higher order differential equations, and state variables then have to be suitably defined to cast the system equations into the state variable form. The situation is similar for distributed systems: by suitably defining the state variables, we can write the system equations as a set of first order (with respect to the time variable) partial differential equations.

For the sake of simplicity let us consider a distributed system involving a single spatial variable $y\varepsilon[0,1]$. Denoting the n-dimensional distributed state vector by $x(y,t)$ and the r-dimensional control vector by $u(y,t)$, the general state equation takes the form

$$\frac{\partial x}{\partial t} = f(x,\frac{\partial x}{\partial y},\ldots,\frac{\partial^k x}{\partial y^k}, u,y,t). \tag{6.2}$$

However, partial differential equations being so much more complex than ordinary differential equations, it is difficult to deduce useful and significant results concerning the solutions of equations in the general form (6.2). Historically, the study of partial differential equations has concentrated on the linear theory and problems that were motivated by physical applications (see [1]). In particular, the following classes of linear partial differential equations have been extensively studied:

(a) elliptic equations,

(b) parabolic equations,

and (c) hyperbolic equations.

Elliptic equations are usually associated with problems in potential theory, e.g., the Laplace and Poisson equations. Parabolic equations are associated with diffusion problems, while hyperbolic equations are associated with problems of wave motion. The boundary

conditions for each type of problems are different, and their solutions exhibit characteristics that are quite different. For example, parabolic and hyperbolic equations are known as evolution equations, and require the specification of some initial data as part of the boundary conditions. In physical applications, time is usually one of the independent variables, hence the name evolution equations.

Control theory deals with dynamical systems and their changes with time, so it is only natural that control problems generally involve systems of the parabolic or the hyperbolic type. The mathematical formulation and analysis of linear control systems governed by parabolic, hyperbolic as well as elliptic equations can be found in the book by Lions [2].

Parabolic Systems

The canonical form for a control system described by a linear parabolic partial differential equation is

$$\frac{\partial x}{\partial t} + Ax = Bu \ , \qquad (y,t)\varepsilon[0,1] \times [0,T] \tag{6.3}$$

where $x(y,t)$ and $u(y,t)$ are the distributed state and control vectors respectively, $B(y,t)$ is a matrix of appropriate dimensions, and A is a linear spatial differential operator which is elliptic (see [2]). The form of (6.3) is the same as that for the linear control system except that the system variables are functions of time and the spatial variable y, and A is now a linear operator instead of being just a matrix as in the lumped case.

The case when A is a second order differential operator is of particular importance in practical applications. A typical example of the parabolic system in this case is given by the equation (in which x

and u are scalar variables)

$$\frac{\partial x}{\partial t} = \frac{\partial}{\partial y}\left[a_1(y,t)\ \frac{\partial x}{\partial y}\right] - a_o(y,t)x + b(y,t)u \tag{6.4}$$

where $a_1(y,t) > 0$ and $a_o(y,t) \geq 0$. The proper boundary conditions for this system are the following conditions at both ends of the spatial boundary,

$$\begin{aligned} c_o x(0,t) + \frac{\partial x}{\partial y}(0,t) &= h(t) \\ c_1 x(1,t) + \frac{\partial x}{\partial y}(1,t) &= p(t) \end{aligned} \tag{6.5}$$

and the initial condition

$$x(y,0) = g_o(y)\ . \tag{6.6}$$

If the system of (6.3) is modified to be of the form

$$\frac{\partial x}{\partial t} = \frac{\partial}{\partial y}\left[a_1(y,t)\ \frac{\partial x}{\partial y}\right] - a_o(y,t)x + b(x,u) \tag{6.7}$$

where b is a nonlinear function of x and u, then it is referred to as a nonlinear (or semi-linear) system of the parabolic type.

Hyperbolic Systems

The canonical form for a control system described by a linear hyperbolic partial differential equation is

$$\frac{\partial^2 x}{\partial t^2} + Ax = Bu\ , \tag{6.8}$$

where x, u are the state and control vectors respectively, B is a matrix and A is a linear elliptic spatial operator, as for the parabolic system (6.3). A typical example of the hyperbolic system when A is a second order operator is given by the following equation

$$\frac{\partial^2 x}{\partial t^2} = \frac{\partial}{\partial y}\left[a_1(y,t)\ \frac{\partial x}{\partial y}\right] - a_o(y,t)x + b(y,t)u \qquad (6.9)$$

where x and u are scalar variables, $a_1(y,t) > 0$ and $a_o(y,t) \geq 0$. The proper boundary conditions for this system are those for the parabolic system (6.4), viz. conditions (6.5) and (6.6), plus the additional initial condition

$$\frac{\partial x}{\partial t}(y,0) = g_1(y) \ . \qquad (6.10)$$

Equations of the form (6.9) are usually associated with harmonic motion, e.g. the vibrating string.

Remarks

The reduction of a higher order partial differential equation to the canonical form (6.2) which is first order with respect to the time variable can usually be performed in more than one way. As an illustration, consider the linear hyperbolic system

$$\frac{\partial^2 x}{\partial t^2} = \frac{\partial^2 x}{\partial y^2} + u \qquad (6.11)$$

with the homogeneous initial conditions

$$x(y,0) = \frac{\partial x}{\partial t}(y,0) = 0 \qquad (6.12)$$

and appropriate boundary conditions.

We shall now reduce the above system to the canonical first order form in two different ways.

(1) The first approach consists of defining new state variables x_1, x_2 by

$$x_1 \equiv x \ , \quad x_2 \equiv \frac{\partial x}{\partial t}$$

Then we can re-write the above system equation (6.11) as

$$\frac{\partial}{\partial t}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \partial^2/\partial y^2 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \ , \tag{6.13}$$

with

$$x_1(y,0) = x_2(y,0) = 0.$$

(2) An alternative approach is to define the state variables

$$x_1 \equiv \frac{\partial x}{\partial y} \ , \quad x_2 \equiv \frac{\partial x}{\partial t} \ .$$

Then the system equation (6.11) can be re-written as

$$\frac{\partial}{\partial t}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 & \partial/\partial y \\ \partial/\partial y & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u \ . \tag{6.14}$$

Again, we see that the initial conditions are

$$x_1(y,0) = x_2(y,0) = 0.$$

Finally, it must be remarked that although we have reduced the hyperbolic system (6.11) to first order forms (with respect to time) (6.13) and (6.14) which are formally identical to the canonical form (6.2) for a parabolic system, the system equations (6.13) and (6.14) are not parabolic because the corresponding spatial differential operator acting on the state vector is not elliptic. This point is discussed by Lions [2] in more precise mathematical language.

6.3 MULTIVARIATE SPLINE FUNCTIONS

So far in this thesis only univariate spline functions have been mentioned; we now consider multivariate splines of more than one independent variable.

Let $S_{h_1}^{\alpha_1}$ and $S_{h_2}^{\alpha_2}$ be spline spaces generated by the bases $\{\psi_1,\ldots,\psi_{m_1}\}$ and $\{\xi_1,\ldots,\xi_{m_2}\}$ respectively. By taking the tensor product of these two spline spaces, we can form the bivariate spline space

$$S_h^\alpha \equiv S_{h_1}^{\alpha_1} \times S_{h_2}^{\alpha_2} , \tag{6.15}$$

where $\alpha \equiv (\alpha_1,\alpha_2)$ and $h \equiv (h_1,h_2)$. S_h^α is the m_1m_2 dimensional linear space consisting of all functions $s(y,t)$ of the form

$$s(y,t) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} q(i,j)\ \psi_i(y)\xi_j(t) \tag{6.16}$$

where each $q(i,j)$ is a real constant. Similarly, multivariate spline spaces of more than two variables can be formed by taking the tensor product of more than two univariate spline spaces.

The approximation properties of univariate splines are summarised in Appendix A. In this section we discuss the approximation properties of multivariate splines; for further details the references [3] and [10] may be consulted.

Consider the closed rectangular region $\Omega \equiv [0,1] \times [0,T]$ in the yt-plane and the linear space $C^\infty(\Omega)$ of infinitely smooth functions defined on Ω. We define the norm $||.||_{2,r}$ on $C^\infty(\Omega)$ as follows:

$$||f||_{2,r} \equiv \left[\int_0^T \int_0^1 \sum_{i,j} (\partial^{i+j}f/\partial y^i \partial t^j)^2 dy\ dt\right]^{\frac{1}{2}} \tag{6.17}$$

for all $f \in C^{\infty}(\Omega)$, where the summation runs over all the non-negative integers i,j such that $0 \leq i+j \leq r$. For the special case $r = 0$, we shall also denote the norm $||.||_{2,0}$ by $||.||_2$.

Let S_h^{α} denote the family of bivariate spline spaces of fixed order $\alpha - 1$ and parametrized by the mesh size h. Then there exists a linear map $L_h: C^{\infty}(\Omega) \to S_h^{\alpha}$ such that for every $f \in C^{\infty}(\Omega)$,

$$||L_h f - f||_{2,r} = \sum_{i=1}^{2} O(h_i^{\alpha_i - r}) \tag{6.18}$$

for $r \leq \min(\alpha_1, \alpha_2)$.

We remark here that the error bound (6.18) still remains valid even when the approximation $L_h f$ is required to satisfy certain boundary conditions. To illustrate, let us consider a smooth function $f(y,t)$ defined on Ω. The error bound of (6.18) can be achieved by choosing L_h to be the usual bicubic interpolation map (see [3]). Furthermore, if the function f satisfies a boundary condition of the type

$$c\, f(0,t) + \frac{\partial f}{\partial y}(0,t) = 0, \tag{6.19}$$

where c is a constant, then it is not hard to verify, using the bicubic interpolatory conditions given in [3], that the above linear boundary condition is also satisfied by the bicubic interpolate.

6.4 THE SP PROCEDURE

We shall now describe the SP procedure for a class of distributed system problems. As mentioned in the introduction we do not attempt to specify the most general class of problems for which the procedure is applicable. In general, whether the procedure can be applied to a

particular problem can be decided simply by inspection. Furthermore, in our description of the solution procedure we restrict the parametrization to multivariate spline spaces, but it is clear that other types of approximation space can also be used.

Problem Formulation

Consider a system described by a nonlinear partial differential equation of the form

$$\frac{\partial x}{\partial t} = f(x, \frac{\partial x}{\partial y}, \ldots, \frac{\partial^k x}{\partial y^k}, u) \tag{6.20}$$

where $x(y,t)$ and $u(y,t)$ are the distributed state and control scalar variables respectively.

The initial condition is assumed to be given by

$$x(y,0) = g_o(y) \tag{6.21}$$

and the boundary conditions are assumed to be an appropriate number of suitable linear combinations of the following conditions:

$$x(0,t) = h_o(t),\ \frac{\partial x}{\partial y}(0,t) = h_1(t), \ldots, \frac{\partial^{k-1} x}{\partial y^{k-1}}(0,t) = h_{k-1}(t)$$

$$x(1,t) = p_o(t),\ \frac{\partial x}{\partial y}(1,t) = p_1(t), \ldots, \frac{\partial^{k-1} x}{\partial y^{k-1}}(1,t) = p_{k-1}(t). \tag{6.22}$$

The problem statement is: find the optimal control $u^*(y,t)$ that minimizes the general cost functional

$$J = \int_0^T \int_0^1 \phi(x, \frac{\partial x}{\partial y}, \ldots, \frac{\partial^k x}{\partial y^k}, u)\, dy\, dt. \tag{6.23}$$

We make the following assumptions on the problem:

(A1) The control variable u(y,t) can be uniquely determined from the system equation (6.20) in terms of x(y,t) and its derivatives:

$$u = g(x, \frac{\partial x}{\partial y}, \ldots, \frac{\partial^k x}{\partial y^k}, \frac{\partial x}{\partial t}). \tag{6.24}$$

(A2) The functions g and ϕ are differentiable in their arguments.

Method of Solution

We first parametrize the state variable x(y,t) as a bivariate spline in the variables y and t. By assumption (A1) it is then possible to derive a corresponding expression for the control u(y.t).

The system equation (6.20) contains partial derivatives of x(y,t) in both variables, up to first order in t and up to order k in y. It is then clear that if a continuous control profile is required, the state variable x(y,t) must be a quadratic or higher order spline in t and a (k+1)-th or higher order spline in y.

Suppose that we take a suitable partition of the spatial interval $[0,1]$ and let $\{\psi_1(y),\ldots,\psi_{m_1}(y)\}$ be a corresponding spline basis, where the degree α_1 of the splines is greater than or equal to k+1. Similarly, we let $\{\xi_1(t),\ldots,\xi_{m_2}(t)\}$ be a spline basis over the time interval $[0,T]$, where the degree α_2 of the splines is greater than or equal to 2.

The expression for the state x(y,t) may then be written

$$x(y,t) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} q(i,j)\psi_i(y)\xi_j(t) , \tag{6.25}$$

where the q(i,j) are the unknown parameters to be optimized. These parameters are not completely free, since the state variable is required to satisfy initial and boundary conditions.

At the initial time t = 0,

$$x(y,0) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} q(i,j)\psi_i(y)\xi_j(0) \tag{6.26}$$

which is a spline function of order α_1 in the spatial variable y.

Since the initial function $g_o(y)$ is not in general a spline function, we cannot expect the initial condition (6.21) to be exactly satisfied by the parametrization (6.25). Instead, we will have to be content with satisfying the initial condition approximately. Thus it will be necessary to fit the spatial modes to $g_o(y)$ in some manner (e.g. performing an interpolation, or finding the least squares fit). However, it is frequently the case that the function $g_o(y)$ is polynomial, in which case no fitting is required provided of course that the order of the polynomial is not higher than the order of the spline space employed.

Suppose that $g_o(y)$ is replaced by the spline function $\tilde{g}_o(y)$,

$$\tilde{g}_o(y) = \sum_{i=1}^{m_1} \beta_i\psi_i(y) \ . \tag{6.27}$$

Using (6.26) and requiring the initial condition to be $\tilde{g}_o(y)$, we have

$$\sum_{i=1}^{m_1} \sum_{j=1}^{m_2} q(i,j)\psi_i(y)\xi_j(0) = \sum_{i=1}^{m_1} \beta_i\psi_i(y) \tag{6.28}$$

Equating coefficients of each $\psi_i(y)$, the following linear system results:

$$\sum_{j=1}^{m_2} q(i,j)\xi_j(0) = \beta_i \ , \qquad i = 1,\ldots,m_1 \tag{6.29}$$

which represents m_1 conditions on the m_1m_2 unknowns q(i,j).

The boundary conditions are treated in a similar manner. Thus for the boundary condition

$$\frac{\partial x}{\partial y}(0,t) = h_1(t) \ , \tag{6.30}$$

we first replace $h_1(t)$, if necessary, by an appropriate spline function which is consistent with the initial condition,

$$\tilde{h}_1(t) = \sum_{j=1}^{m_2} \gamma_j \xi_j(t) \ , \tag{6.31}$$

and using (6.25), we obtain the following linear conditions:

$$\sum_{i=1}^{m_1} q(i,j)\psi_i'(y) = \gamma_j \ , \qquad j = 1,\ldots,m_2 \tag{6.32}$$

where ψ_i' denotes the derivative of ψ_i with respect to y.

The above procedure can be repeated for all the specified boundary conditions. In this way, the initial and boundary conditions are reduced to linear conditions on the unknowns q(i,j).

When the appropriate parametric expressions are used in the expression (6.23) for the cost functional J, we can write the cost as a function $\tilde{J}(q)$ of the unknown parameter matrix q. The function $\tilde{J}$ is to be minimized with respect to the parameter matrix q subject to the linear constraints imposed by the initial and boundary conditions.

This optimization problem involving equality constraints may be solved using the quadratically convergent algorithm of Goldfarb and Lapidus [4]. The gradient of $\tilde{J}$ with respect to q required by the algorithm can be evaluated simply by differentiating the right-hand side of (6.23) under the double integral,

$$\frac{\partial \tilde{J}}{\partial q} = \int_0^T \int_0^1 \frac{\partial \phi}{\partial q} \left(x, \frac{\partial x}{\partial y}, \ldots, \frac{\partial^k x}{\partial y^k}, u\right) dy\, dt \tag{6.33}$$

and applying the chain rule of differentiation. It is understood that the gradient of $\tilde{J}$ is to be evaluated with respect to each element q(i,j) in the above expression.

Remarks:

Our procedure requires that the approximate initial condition be consistent (or compatible) with the approximate boundary conditions; i.e. they should match at the corners of the rectangular domain $[0,T] \times [0,1]$ so that x(y,t) would remain smooth on the boundary. Thus, for instance we require that

$$\tilde{g}_o'(0) = \tilde{h}_1(0). \tag{6.34}$$

Therefore, the spline approximations for the initial and boundary conditions must be chosen with this consistency condition in mind.

Consequently, the constraints on q(i,j) that we obtain from the initial conditions and boundary conditions separately cannot be completely independent. For example, we have m_1 constraints in (6.29) and m_2 constraints in (6.32), but owing to the consistency condition (6.34), only $m_1 + m_2 - 1$ constraints are independent. In this case, it would be desirable to remove the superfluous constraint to reduce the computational load. However, we must also remark that the Goldfarb-Lapidus algorithm is also applicable to a set of linearly dependent constraints.

Finally, we note that the linearity of the constraints on q(i,j) depends only on the linearity of the corresponding initial and boundary conditions, and not on the linearity of the system equation, thereby permitting the use of the efficient Goldfarb-Lapidus algorithm even when the system is nonlinear.

Numerical Examples

We now apply the SP procedure to solve two specific distributed parameter system problems. One of these problems involves a linear parabolic system while the other involves a nonlinear system of the hyperbolic type. In both cases a bicubic spline parametrization of the state variable is employed.

Example 1

This problem is taken from Chaudhuri [8], but the initial conditions have been modified to make them consistent with the boundary conditions.

For the nonlinear system

$$\frac{\partial^2 x}{\partial t^2} = \frac{\partial^2 x}{\partial y^2} - x^3 + u \ , \quad (y,t) \in [0,1.5] \times [0,.4] , \tag{6.35}$$

with the initial conditions

$$x(y,0) = 1 + 2y^2 - \frac{8}{9} y^3 \ , \quad \frac{\partial x}{\partial t}(y,0) = 4y^2 - \frac{16}{9} y^3 \ , \tag{6.36}$$

and the boundary conditions

$$\frac{\partial x}{\partial y}(0,t) = 0 \ , \quad \frac{\partial x}{\partial y}(1.5,t) = 0 \ , \tag{6.37}$$

find the control $u^*(y,t)$ that minimizes the cost

$$J = \frac{1}{2} \int_0^{.4} \int_0^{1.5} \left[2\left(\frac{\partial x}{\partial t}\right)^2 - .5\left(\frac{\partial x}{\partial y}\right)^2 + u^2 \right] dy \, dt \ . \tag{6.38}$$

For the above problem we take the uniform partition of the spatial interval $[0,1.5]$ with M sections and the uniform partition of the time interval $[0,.4]$ with N sections, and adopt a bicubic spline parametrization of the state variable $x(y,t)$:

$$x(y,t) = \sum_{i=1}^{M+3} \sum_{j=1}^{N+3} q(i,j) \psi_i(y) \xi_j(t) \tag{6.39}$$

where $\psi_i(y)$ and $\xi_j(t)$ are cubic B-spline basis functions in the variables y and t respectively.

The system equation can easily be solved for the control u(y,t), and the parametric expression for the control is

$$u(y,t) = x^3 + \sum_{i=1}^{M+3} \sum_{j=1}^{N+3} q(i,j)\left[\psi_i(y)\ddot{\xi}_j(t) - \psi_i''(y)\xi_j(t)\right] \qquad (6.40)$$

where $\ddot{\xi}_j(t)$ denotes the second-order derivative of ξ with respect to t, and $\psi_i''(y)$ denotes the second-order derivative of ψ_i with respect to y. It is clear that the above control expression is continuous.

Next we consider the initial conditions (6.36) and the boundary conditions (6.37). Since the initial conditions are cubic polynomials consistent with the homogeneous boundary conditons, no modification is necessary.

For the initial conditions (6.35) we can obtain, using (6.39) that

$$\sum_{i=1}^{M+3} \sum_{j=1}^{N+3} q(i,j)\psi_i(y)\xi_j(0) = 1 + 2y^2 - \frac{8}{9}y^3 ,$$

$$\sum_{i=1}^{M+3} \sum_{j=1}^{N+3} q(i,j)\psi_i(y)\xi_j(0) = 4y^2 - \frac{16}{9}y^3 . \qquad (6.41)$$

The cubic polynomials can be expressed as a linear combination of cubic B-splines in the spatial interval $[0,1.5]$; i.e. we can determine the constants β_i and γ_i such that

$$1 + 2y^2 - \frac{8}{9}y^3 = \sum_{i=1}^{M+3} \beta_i\psi_i(y) ,$$

$$4y^2 - \frac{16}{9}y^3 = \sum_{i=1}^{M+3} \gamma_i\psi_i(y) , \qquad (6.42)$$

for $y \in [0,1.5]$. A general procedure for determining the representation of polynomials by B-splines is given in the paper of Marsden [9].

From (6.41), (6.42) and the numerical properties of cubic B-splines, the following linear conditions on q are obtained:

$$\begin{aligned} q(i,1) + 4q(i,2) + q(i,3) &= 6\beta_i \ , \\ - q(i,1) + q(i,3) &= 2h\gamma_i \ , \end{aligned} \qquad i = 1 \ , \ldots , \ M+3 \qquad (6.43)$$

where $h \equiv .4/N$ is the mesh size along the t-axis.

For the boundary conditions (6.36) we obtain using (6.39) that

$$\begin{aligned} &\sum_{i=1}^{M+3} \sum_{j=1}^{N+3} q(i,j)\psi_i'(0)\xi_j(t) = 0 \ , \\ &\sum_{i=1}^{M+3} \sum_{j=1}^{N+3} q(i,j)\psi_i'(1.5)\xi_j(t) = 0 \ , \end{aligned} \qquad (6.44)$$

from which the following linear equations are obtained:

$$\begin{aligned} q(1,j) &= q(3,j) \ , \\ q(M+1,j) &= q(M+3,j) \ . \end{aligned} \qquad j = 1,\ldots,N+3 \qquad (6.45)$$

Not all the constraints on q(i,j) given in (6.43) and (6.45) are independent. In fact, the consistency conditions require that $\beta_1 = \beta_3$, $\beta_{M+1} = \beta_{M+3}$, $\gamma_1 = \gamma_3$ and $\gamma_{M+1} = \gamma_{M+3}$. Hence, given the constraints in (6.45), the constraints in (6.43) corresponding to i = 1 and i = 3 are identical. Similarly, the constraints in (6.43) corresponding to i = M + 1 and i = M + 3 are also identical.

Employing the 8 × 8 Gauss-Legendre quadrature scheme to evaluate the double integrals, the problem was solved for several values of M and N. The minimum cost obtained for each case is included in the following Table 6.1

TABLE 6.1 MINIMUM COST $J(\bar{x})$

M \ N	1	2	3
1	1.737	1.484	1.472
2	1.353	1.087	1.072
3	1.338	1.073	1.058

The cost figures of Table 6.1 agree with those obtained using the 10 × 10 Gauss-Legendre quadrature scheme over each sub-rectangle in the uniform mesh. Thus for this problem we obtained results of reasonable accuracy by using the coarser quadrature scheme. The control and state profiles are summarised below for the case M = N = 3.

TABLE 6.2 CONTROL PROFILE (M = N = 3)

t	$\bar{u}(0,t)$	$\bar{u}(0.5,t)$	$\bar{u}(1.0,t)$	$\bar{u}(1.5,t)$
0.00	-.6662E0	-.2315E0	-.2095E0	-.4684E0
0.05	-.8367E0	-.2083E0	.2850E0	.9487E0
0.10	-.8908E0	-.1445E0	.4821E0	.1115E1
0.15	-.9117E0	-.7535E-1	.6283E0	.9204E0
0.20	-.9824E0	-.4002E-1	.1008E1	.1493E1
0.25	-.9253E0	.4730E-1	.8168E0	.8513E0
0.30	-.9110E0	.9503E-1	.8213E0	.3394E0
0.35	-.8426E0	.1291E0	.7226E0	-.4424E-1
0.40	-.6207E0	.1663E0	.2644E0	-.3236E0

TABLE 6.3 STATE PROFILE (M = N = 3)

t	$\bar{x}(0,t)$	$\bar{x}(0.5,t)$	$\bar{x}(1.0,t)$	$\bar{x}(1.5,t)$
0.00	.1000E1	.1389E1	.2111E1	.2500E1
0.05	.1003E1	.1426E1	.2208E1	.2624E1
0.10	.1012E1	.1458E1	.2275E1	.2695E1
0.15	.1028E1	.1487E1	.2310E1	.2710E1
0.20	.1051E1	.1510E1	.2310E1	.2667E1
0.25	.1081E1	.1529E1	.2277E1	.2573E1
0.30	.1116E1	.1541E1	.2210E1	.2433E1
0.35	.1158E1	.1548E1	.2112E1	.2255E1
0.40	.1205E1	.1546E1	.1986E1	.2046E1

Example 2

This problem which is taken from [6] involves a linear parabolic control system described by

$$\frac{\partial x}{\partial t} = \frac{\partial^2 x}{\partial y^2} + u \ , \qquad (y,t) \in [0,1] \times [0,1] .$$

with the initial condition

$$x(y,0) = 1 - 6y^2 + 4y^3 \ ,$$

and the boundary conditions

$$\frac{\partial x}{\partial y}(0,t) = 0 \ , \quad \frac{\partial x}{\partial y}(1,t) = 0 \ .$$

The problem here is to find the control $u^*(y,t)$ that minimizes the cost functional

$$J = \frac{1}{2} \int_0^1 \int_0^1 [x^2 + u^2] \, dy \, dt \ .$$

As for Example 1, a bicubic spline parametrization of $x(y,t)$ was adopted for this problem. Numerical results were obtained for several values of M and N (keeping M=N in each case) and are presented in Tables 6.4, 6.5 and 6.6. The cost functional was evaluated using the 4×4 Gauss-Legendre quadrature scheme over each sub-rectangle of the mesh, and is therefore exact to within machine accuracy.

TABLE 6.4 MINIMUM COST

M=N	1	2	3	4	5
$J(\bar{x})$	0.32976	0.10615	0.03766	0.02206	0.01751

Tables 6.5 and 6.6 below summarize the final control and state profiles obtained for the case M = N = 3. The control and state obtained are both anti-symmetric about the line $y = 0.5$; i.e.,

$$\bar{u}(1-y,t) = -\bar{u}(y,t) \text{ and } \bar{x}(1-y,t) = -\bar{x}(y,t)$$

for $y \in [0,.5]$ and $t \in [0,1]$. Note that this means that $\bar{u}(.5,t) = 0$ and $\bar{x}(.5,t) = 0$. This feature of the solution is also obtained for the other values of M and N used.

TABLE 6.5 CONTROL PROFILE (M = N = 3)

t	$\bar{u}(0,t)$	$\bar{u}(.1,t)$	$\bar{u}(.2,t)$	$\bar{u}(.3,t)$	$\bar{u}(.4,t)$
0.0	.3426E1	.1535E1	.4924E0	.4511E-1	-.5511E-1
0.1	-.7387E-1	-.4791E0	-.4789E0	-.1808E0	-.4777E-1
0.2	-.3759E0	-.3625E0	-.2241E0	.1998E-1	.5872E-1
0.3	.3181E0	.2571E0	.2038E0	.1690E0	.9217E-1
0.4	.2936E0	.1833E0	.8724E-1	.4210E-2	-.1451E-1
0.5	.4500E-1	-.5681E-2	-.3690E-1	-.5700E-1	-.3760E-1
0.6	-.1542E0	-.1245E0	-.8093E-1	-.3063E-1	-.7691E-2
0.7	-.1208E0	-.7374E-1	-.2825E-1	.1770E-1	.1952E-1
0.8	.4846E-1	.4463E-1	.3223E-1	.1690E-1	.6455E-2
0.9	.1019E0	.6816E-1	.3269E-1	-.4649E-2	-.1015E-1
1.0	-.1928E0	-.1415E0	-.7129E-1	.2578E-2	.1820E-1

TABLE 6.6 STATE PROFILE (M = N = 3)

t	$\bar{x}(0,t)$	$\bar{x}(.1,t)$	$\bar{x}(.2,t)$	$\bar{x}(.3,t)$	$\bar{x}(.4,t)$
0.0	.1000E1	.9440E0	.7920E0	.5680E0	.2960E0
0.1	.3754E0	.3559E0	.3018E0	.2195E0	.1156E0
0.2	.1105E0	.1056E0	.9116E-1	.6783E-1	.3636E-1
0.3	.4659E-1	.4428E-1	.3778E-1	.2769E-1	.1467E-1
0.4	.3322E-1	.3113E-1	.2565E-1	.1795E-1	.9171E-2
0.5	.1540E-1	.1438E-1	.1174E-1	.8111E-2	.4101E-2
0.6	-.2642E-3	-.1293E-3	.1384E-3	.3330E-3	.2694E-3
0.7	-.6242E-2	-.5690E-2	-.4359E-2	-.2733E-2	-.1263E-2
0.8	-.1748E-2	-.1661E-2	-.1417E-2	-.1038E-2	-.5499E-3
0.9	.3549E-2	.3203E-2	.2385E-2	.1424E-2	.6248E-3
1.0	-.9191E-3	-.6598E-3	-.1264E-3	.3139E-3	.3295E-3

6.5 ERROR ANALYSIS

We now consider the convergence of the SP solution over spline approximation spaces as the mesh size h goes to zero. In the previous chapter we reformulated the lumped problem involving a linear system in the phase-variable form and a quadratic performance index as a problem in variational calculus with a cost functional that is elliptic under the usual assumptions on the linear quadratic problem. Then, by using known error bounds for the Ritz (or Rayleigh-Ritz) solution to the variational calculus problem over spline approximation spaces, we were able to deduce the corresponding results for the SP solution.

We might ask whether error bounds for the distributed linear quadratic problem can be obtained by the same method. Unfortunately, it appears that we cannot employ this approach in the distributed case. The reason is this: when we cast the given control problem as a variational calculus problem, the cost functional turns out to be not necessarily elliptic; or, what amounts to the same thing, the Euler equation for the problem might not be elliptic. We shall illustrate this point by a simple example.

Example

Let us consider the following parabolic system

$$\frac{\partial x}{\partial t} = \frac{\partial^2 x}{\partial y^2} + u \ , \qquad (y,t) \ \epsilon [0,1] \times [0,T]$$

with given initial and boundary conditions

$$x(y,0) = g_o(y) \ , \qquad \frac{\partial x}{\partial y}(0,t) = \frac{\partial x}{\partial y}(1,t) = 0,$$

and suppose that we wish to find the control $u^*(y,t)$ that minimizes the cost functional

$$J = \frac{1}{2} \int_0^T \int_0^1 (x^2 + u^2) \, dy \, dt \ .$$

The necessary conditions of optimality for this problem are known to be given by (see [5]):

$$\frac{\partial x^*}{\partial t} - \frac{\partial^2 x^*}{\partial y^2} - u^* = 0 ,$$

$$\frac{\partial \lambda^*}{\partial t} + \frac{\partial^2 \lambda^*}{\partial y^2} + x^* = 0 ,$$

$$u^* + \lambda^* = 0 ,$$

and the side conditions

$$x^*(y,0) = g_o(y) , \qquad \frac{\partial x^*}{\partial y}(0,t) = \frac{\partial x^*}{\partial y}(1,t) = 0 ,$$

$$\lambda^*(y,T) = 0 , \qquad \frac{\partial \lambda^*}{\partial y}(0,t) = \frac{\partial \lambda^*}{\partial y}(1,t) = 0 .$$

From the above optimality conditions, we find that the associated Euler equation is

$$\frac{\partial^4 x^*}{\partial y^4} - \frac{\partial^2 x^*}{\partial t^2} + x^* = 0 ,$$

which is certainly not elliptic.

Hence, we cannot apply known results concerning the error bounds of the Ritz solution to our distributed control problem, as these results refer only to elliptic equations. What we have done instead is to adopt an approach similar to that of Chapter 3 to derive error bounds for the SP approximation. While the error bounds obtained in this manner may not be optimal for the state $\bar{x}^h$, they should be optimal for $\bar{u}^h$, and consequently for the cost $J(\bar{x}^h)$ too.

We outline below the derivation of error bounds for a second order linear parabolic system and a quadratic performance index. The method of derivation should also be applicable to other similar problems, for instance, a second order linear hyperbolic system with a quadratic performance index.

Problem Statement

Let us consider the following problem: given the linear dynamical system

$$\frac{\partial x}{\partial t} = \frac{\partial}{\partial y}\left[a(y,t)\ \frac{\partial x}{\partial y}\right] + b(y,t)\ u\ , \tag{6.46}$$

the initial condition

$$x(y,0) = g_o(y)\ , \tag{6.47}$$

and the boundary conditions

$$\begin{aligned} c_o x(0,t) + \frac{\partial x}{\partial y}(0,t) &= h(t)\ , \\ c_1 x(1,t) + \frac{\partial x}{\partial y}(1,t) &= p(t)\ , \end{aligned} \tag{6.48}$$

find the control $u^*(y,t)$ which minimizes the cost functional

$$J = \frac{1}{2}\int_0^T \int_0^1 \left[Q(y,t)x^2 + R(y,t)u^2\right] dy\ dt, \tag{6.49}$$

where x and u are scalar variables. Here c_o and c_1 are constants, and the functions $a(y,t)$, $b(y,t)$, $Q(y,t)$, $R(y,t)$ are assumed to be arbitrarily smooth within the domain $[0,1] \times [0,T]$ of the problem; also, we assume that $b(y,t) \neq 0$, $Q(y,t) \geq 0$ and $R(y,t) > 0$. The functions $g_o(y)$, $h(t)$ and $p(t)$ are assumed to be polynomials satisfying the consistency conditions. Finally, we assume that

$$\left|\left|x - x^*\right|\right|_2 \leq K\left|\left|u - u^*\right|\right|_2 \tag{6.50}$$

for some constant $K > 0$ and all (x,u) satisfying the state equation (6.46) and the side conditions (6.47), (6.48).

Remarks

(i) Suppose the functional $\left|\left|.\right|\right|_R$ is defined by

$$\left|\left|u\right|\right|_R \equiv \left[\int_0^T \int_0^1 R(y,t)\ u^2(y,t)\ dy\ dt\right]^{\frac{1}{2}}\ ,$$

and that $||.||_Q$ is similarly defined.

Since R is assumed to be positive, we can find positive constants γ_1 and γ_2 such that

$$\gamma_1||u||_2^2 \leq ||u||_R^2 \leq \gamma_2||u||_2^2 . \tag{6.51}$$

Similarly, since Q is non-negative, we can find a positive constant γ_3 such that

$$||x||_Q^2 \leq \gamma_3||x||_2^2 . \tag{6.52}$$

(ii) The necessary conditions of optimality for the above problem are given by (see [5])

$$\frac{\partial\lambda_1^*}{\partial t}(y,t) + \frac{\partial}{\partial y}\left[a(y,t)\frac{\partial\lambda_1^*}{\partial y}(y,t)\right] + Q(y,t)\,x^*(y,t) = 0,$$

$$R(y,t)\,u^*(y,t) + b(y,t)\,\lambda_1^*(y,t) = 0 ,$$

for all $(y,t) \in [0,1] \times [0,T]$, and the side conditions

$$\lambda_1^*(y,T) = 0, \qquad \lambda_1^*(y,0) + \lambda_2^*(y) = 0,$$

$$\lambda_3^*(t) + a(0,t)\,\lambda_1^*(0,t) = 0 ,$$

$$c_o\lambda_3^*(t) - a(0,t)\frac{\partial\lambda_1^*}{\partial y}(0,t) = 0 ,$$

$$\lambda_4^*(t) - a(1,t)\,\lambda_1^*(1,t) = 0 ,$$

$$c_1\lambda_4^*(t) + a(1,t)\frac{\partial\lambda_1^*}{\partial y}(1,t) = 0 ,$$

where the optimal variables $x^*(y,t)$ and $u^*(y,t)$ also satisfy equations (6.46) - (6.48).

Using these optimality conditions, it is a straightforward matter to verify that

$$J(x) = J(x^*) + \frac{1}{2}(||x-x^*||_Q^2 + ||u-u^*||_R^2) \tag{6.53}$$

for all (x,u) satisfying (6.46) – (6.48).

(iii) The smoothness assumptions on the problem imply that the optimal variables x* and u* are also smooth in the region $[0,1] \times [0,T]$.

(iv) The property (6.50) follows from the existence of a family of bounded linear transformations for the system (6.46) which plays a similar role to that of the transition matrix for a lumped linear system. A proper appreciation of this point requires some knowledge of the theory of one-parameter semigroups of linear operators (see [11]), so we shall be content with assuming the property (6.50).

Derivation of Error Bounds

Consider the SP solution over the bivariate spline approximation space S_h^α, where it is assumed that the order of S_h^α is sufficiently high for the side conditions (6.47) and (6.48) to be satisfied exactly.

Now, we know that there exists $x_s^h \in S_h^\alpha$ such that (6.47), (6.48) are satisfied, and that

$$||x_s^h - x^*||_2 \leq O(h^\alpha). \tag{6.54}$$

It follows that

$$||u_s^h - u^*||_2 = ||\frac{\partial}{\partial t}(x_s^h - x^*) - \frac{\partial}{\partial y}[a\frac{\partial}{\partial y}(x_s^h - x^*)]||_2 \leq O(h^{\alpha-2}). \tag{6.55}$$

By definition of the SP procedure,

$$J(\bar{x}^h) \leq J(x_s^h). \tag{6.56}$$

Applying identity (6.53) to the above inequality, we obtain that

$$||\bar{x}^h - x^*||_Q^2 + ||\bar{u}^h - u^*||_R^2 \leq ||x_s^h - x^*||_Q^2 + ||u_s^h - u^*||_R^2 , \tag{6.57}$$

and using (6.51), (6.52) we deduce that

$$\begin{aligned} ||\bar{u}^h - u^*||_R^2 &\leq \gamma_3 ||x_s^h - x^*||_2^2 + \gamma_2 ||u_s^h - u^*||_2^2 \\ &\leq O(h^{2\alpha}) + O(h^{2(\alpha-2)}) \\ &= O(h^{2(\alpha-2)}). \end{aligned} \tag{6.58}$$

Hence,

$$||\bar{u}^h - u^*||_R \leq O(h^{\alpha-2}), \tag{6.59}$$

and by (6.51) this is equivalent to

$$||\bar{u}^h - u^*||_2 \leq O(h^{\alpha-2}). \tag{6.60}$$

Using (6.50), it follows from (6.60) that

$$||\bar{x}^h - x^*||_2 \leq O(h^{\alpha-2}). \tag{6.61}$$

Finally, by applying (6.54) and (6.55) to (6.57), we obtain that

$$0 \leq J(\bar{x}^h) - J(x^*) \leq O(h^{2(\alpha-2)}). \tag{6.62}$$

6.6 CONCLUSIONS

We have examined the SP procedure for solving a general class of distributed system problems involving distributed controls. In particular, we have described the procedure in conjunction with a bivariate spline approximation space. Using this method, numerical results for two specific problems have been obtained.

Finally, an error analysis of the procedure has been carried out for a class of linear distributed system problems involving quadratic performance indices. We noted that the error bound obtained for the state variable convergence was likely to be only sub-optimal.

REFERENCES

[1] R. Courant and D. Hilbert: Methods of Mathematical Physics, Vol.II, Interscience, N.Y. (1953).

[2] J. L. Lions: Optimal Control of Systems Governed by Partial Differential Equations, Springer-Verlag, N.Y. (1971).

[3] M. H. Schultz: Spline Analysis, Prentice-Hall Inc., N.J. (1973).

[4] D. Goldfarb and L. Lapidus: Conjugate Gradient Method for Nonlinear Programming Problems with Linear Constraints, Ind. Eng. Chem. Fundamentals, Vol.7 (1968) pp.142-151.

[5] W. E. Bosarge, Jr., O. G. Johnson and C. L. Smith: A Direct Method Approximation to the Linear Parabolic Regulator Problem over Multivariate Spline Bases, SIAM J. Numer. Anal., Vol. 10 (1973) pp.35-49.

[6] W. E. Bosarge, Jr. and C. L. Smith: Numerical Properties of a Multivariate Ritz-Trefftz Method, IBM J. Res. Develop., Vol.16 (1972) pp.393-400.

[7] R. S. McKnight and W. E. Bosarge, Jr.: The Ritz-Galerkin Procedure for Parabolic Control Problems, SIAM J. Control, Vol.11 (1973) pp.510-524.

[8] S. P. Chaudhuri: Optimal Control Computational Techniques for a Class of Non-linear Distributed Parameter Systems, Int. J. Control, Vol.15 (1972) pp.419-432.

[9] M. J. Marsden: An Identity for Spline Functions with Applications to Variation-Diminishing Spline Approximation, J. Approx. Theory, Vol.3 (1970) pp.7-49.

[10] M. H. Schultz: Approximation Theory of Multivariate Spline Functions in Sobolev Spaces, SIAM J. Numer. Anal., Vol.6 (1969) pp.570-582.

[11] A. V. Balakrishnan: Introduction to Optimization Theory in a Hilbert Space, Springer-Verlag, N.Y. (1971).

CHAPTER 7

GENERAL CONCLUSIONS

The purpose of this concluding chapter is two-fold: to summarise the contributions that this thesis makes towards the computational theory of optimal control, and to indicate possible directions in which the present work can be extended.

In this project we have focused attention on two specific parametrization procedures, viz. the CP and SP procedures. The broad objective of this research has been to investigate the application of spline functions in conjunction with these parametrization procedures. This has been achieved by

(a) examining the computational aspects of each parametrization technique, and

(b) deriving the relevant error bounds for the approximate solution in terms of the mesh size.

The highlights of this thesis are now reviewed.

A series of articles by Bosarge et.al. [1] - [4] examined the convergence properties of the Ritz-Trefftz and Ritz-Galerkin solutions; more specifically, error bounds were derived in terms of the mesh size in the case of piecewise polynomial approximation spaces. In Chapter 3 of this thesis we followed up the work of Bosarge and his co-workers by conducting a similar analysis on the CP procedure. The error bounds obtained provide useful indications of the accuracy of the CP solution.

In Chapter 4 the SP procedure was proposed as a viable alternative to the CP procedure. The SP procedure is based on treating one or more state variables as the independent variables, and for certain classes

of problems the method is computationally more efficient than the CP method.

In the case of spline approximation spaces, we obtained error bounds for the SP solution in Chapter 5 for a class of optimal control problems. The results were obtained by relating the SP procedure to the classical Rayleigh-Ritz procedure for solving problems of variational calculus.

We then extended the SP technique to solve problems in distributed parameter systems involving distributed controls. An error analysis was also carried out for a class of control problems, but it is felt that the error bounds obtained are by no means optimal. It would be interesting, though, if improved bounds could be found.

Finally, we remark that throughout this thesis, we have only employed splines with fixed knots. (As a matter of fact, all our computations have been performed with evenly spaced knots). However, the knot locations can be treated as additional variables to be optimized. The importance of using optimal knots in the approximation of functions by splines has been emphasized by Burchard [5], whose paper also contains some results on the convergence of nonlinear (i.e. variable knots) spline approximation. It might be worthwhile to pursue an investigation of the application of the nonlinear spline parametrization to optimal control problems, particularly those with non-smooth solutions.

REFERENCES

[1] W. E. Bosarge, Jr. and O. G. Johnson: Error Bounds of High Order Accuracy for the State Regulator Problem via Piecewise Polynomial Approximations, SIAM J. Control, Vol.9 (1971), pp.15-28.

[2] W. E. Bosarge, Jr., O. G. Johnson, R. S. McKnight and W. P. Timlake: The Ritz-Galerkin Procedure for Nonlinear Control Problems, SIAM J. Numer. Anal., Vol.10 (1973), pp.94-111.

[3] W. E. Bosarge, Jr., O. G. Johnson and C. L. Smith: A Direct Method Approximation to the Linear Parabolic Regulator Problem over Multivariate Spline Bases, SIAM J. Numer. Anal., Vol.10 (1973) pp.35-49.

[4] R. S. McKnight and W. E. Bosarge, Jr.: The Ritz-Galerkin Procedure for Parabolic Control Problems, SIAM J. Control, Vol.11 (1973) pp.510-524.

[5] H. G. Burchard: Splines (with Optimal Knots) are Better, Applicable Analysis, Vol.3 (1974), pp.309-319.

APPENDIX A

MATHEMATICAL BACKGROUND

Introduction

This appendix contains a collection of basic mathematical definitions and results in functional analysis and approximation theory which constitutes roughly the mathematical background essential for a proper understanding of Chapter 3. At the same time it also serves a dual purpose as a convenient reference for some mathematical notations. General references for the material in this appendix are [1], [2] and [3].

Normed Linear Space

Definition:

A norm on a linear space X is a function $||.||: X \to E$ such that

(a) $||\alpha x|| = |\alpha| . ||x||$ for all $\alpha \in E$, $x \in X$,

(b) $||x+y|| \leq ||x|| + ||y||$ for all $x, y \in X$,

(c) $||x|| \geq 0$ for all $x \in X$,

(d) $||x|| = 0$ if and only if $x = 0$.

Example:

Let $C[0,T]$ denote the collection of functions which are continuous in the interval $[0,T]$. Two examples of norms defined on $C[0,T]$ are:

(a) $||f|| \equiv \max\{|f(x)| \mid x \in [0,T]\}$

(b) $||f|| \equiv \left[\int_0^T f^2(x)\, dx\right]^{\frac{1}{2}}$.

Definition:

A linear space with a given norm is called a normed linear space (NLS). If every Cauchy sequence in a NLS X converges in X, then it is said to be complete, and is called a Banach space.

Example:

Let $L_p[0,T]$, $p \geq 1$, denote the collection of all real-valued functions defined on $[0,T]$ which are pth-integrable (in the Lebesgue sense); i.e.,

$$\int_0^T |f(t)|^p \, dt < \infty.$$

Then $L_p[0,T]$ is a Banach space with the norm

$$||f||_p = \left[\int_0^T |f(t)|^p \, dt\right]^{1/p}.$$

Example:

The Sobolev space $W_p^\alpha[0,T]$ is the set of all real-valued functions f(t) whose αth-derivatives belong to $L_p[0,T]$; i.e.,

$$W_p^\alpha[0,T] \equiv \{f \mid \int_0^T (\sum_{i=0}^{\alpha} |d^i f/dt^i|^p) dt < \infty\}.$$

With the norm

$$||f||_{p,\alpha} \equiv \left[\int_0^T (\sum_{i=0}^{\alpha} |d^i f/dt^i|^p) \, dt\right]^{1/p},$$

$W_p^\alpha[0,T]$ is a Banach space. If $\alpha = 0$, the norm is identical to the L_p-norm, so $||f||_p \equiv ||f||_{p,0}$.

The space $\{W_p^\alpha[0,T], E^n\}$ is defined as the set of all n-vector valued functions $f = (f_1,\ldots,f_n)$ defined on $[0,T]$ such that

$$||f||_{p,\alpha} \equiv \left[\sum_{i=1}^{n} ||f_i||_{p,\alpha}^p\right]^{1/p} < \infty.$$

Existence of Best Approximation

Definition:

A NLS X is uniformly convex if for every $\varepsilon > 0$ there exists $\delta > 0$ such that $||x-y|| < \varepsilon$ whenever $||x|| = ||y|| = 1$ and $||\frac{x+y}{2}|| > 1 - \delta$, where $x,y \in X$.

Theorem:

If X is a uniformly convex Banach space and W is a closed convex subset of X, then for each $x \in X$ there exists a unique $w^* \in W$ such that

$$||x-w^*|| \leq ||x-w|| \qquad \text{for all } w \in W.$$

Approximation Property of Splines

We now cite an important property of spline functions. Further details may be found in [4], [5] and the references therein.

Theorem:

Suppose S_h^α (where α is an integer and $\alpha \geq 1$) is a space of spline functions of order $\alpha - 1$ parametrized by the mesh size h. Then there exists a linear operator $L_h : PC^\alpha[0,T] \to S_h^\alpha$ such that for every $f \in PC^\alpha[0,T]$,

$$||D^i(L_h f - f)||_2 = O(h^{\alpha-i})$$

for all integer i, where $0 \leq i \leq \alpha$.

REFERENCES

[1] A. E. Taylor: Introduction to Functional Analysis, John Wiley, N.Y. (1958).

[2] D. G. Luenberger: Optimization by Vector Space Methods, John Wiley, N.Y. (1968).

[3] E. W. Cheney: Introduction to Approximation Theory, McGraw-Hill, N.Y. (1966).

[4] L. L. Schumaker: Approximation by Splines, in Theory and Applications of Spline Functions, ed. T.N.E. Greville, Academic Press, N.Y. (1969).

[5] M. H. Schultz: Spline Analysis, Prentice-Hall Inc., N.J. (1973).

APPENDIX B

A SUMMARY OF PARAMETRIZATION PROCEDURES

Introduction

We stated in Chapter 1 that there exist several ways of discretizing a given optimal control problem. Each of these methods involve restricting one or more of the control, state and co-state variables to finite-dimensional approximation spaces. In this appendix we summarize the main features of the following parametrization techniques:

(1) control parametrization,

(2) Ritz-Trefftz,

(3) trajectory approximation, and

(4) Ritz-Galerkin.

Problem Statement and Optimality Conditions

We assume here the following continuous-time optimal control problem: for the given system

$$\dot{x} = f(x,u,t) , \quad x(0) = x_o \tag{B.1}$$

find the optimal control u^* which minimizes the cost functional

$$J = \int_0^T \phi(x,u,t)\, dt, \tag{B.2}$$

where the final time T is taken to be fixed. We assume that the state x is an n-dimensional vector and that the control u is an r-dimensional vector.

The Hamiltonian for the above problem is defined by

$$H \equiv \phi + \langle \lambda, f \rangle \ , \tag{B.3}$$

where λ is the n-dimensional co-state vector given by the equation

$$\dot{\lambda} + \frac{\partial H}{\partial x} = 0 \ , \qquad \lambda(T) = 0 \ . \tag{B.4}$$

It is well known that the necessary condition for optimality is

$$\frac{\partial H}{\partial u} = 0 \ . \tag{B.5}$$

Summarizing, the optimal triple (x^*,u^*,λ^*) satisfies the following conditions simultaneously:

$$\begin{aligned} &\dot{x}^* = f(x^*,u^*,t) \ , && x^*(0) = x_o \ , \\ &\dot{\lambda}^* + \frac{\partial H}{\partial x}(x^*,u^*,\lambda^*,t) = 0 \ , && \lambda^*(T) = 0 \ , \\ &\frac{\partial H}{\partial u}(x^*,u^*,\lambda^*,t) = 0 \ . \end{aligned} \tag{B.6}$$

An alternative characterization of the optimal triple (x^*,u^*,λ^*) involves the Lagrangian functional $L(x,u,\lambda)$ which is defined by

$$L(x,u,\lambda) \equiv J + \int_0^T \langle -\dot{x} + f, \lambda \rangle \, dt \ . \tag{B.7}$$

The theory of multipliers tells us that (x^*,u^*,λ^*) is a saddle point of the Lagrangian (see [1]),

$$L(x^*,u^*,\lambda^*) = \sup_{\lambda \in A(\lambda)} \left[\inf \{ L(x,u,\lambda) \mid u \in A(u), x \in A(x) \} \right], \tag{B.8}$$

where $A(u)$, $A(x)$ and $A(\lambda)$ denote admissible spaces for the control, state and co-state respectively, $A(\lambda)$ being the dual space of $A(x)$.

The Parametrization

Let $S(u)$, $S(x)$ and $S(\lambda)$ be finite-dimensional subspaces of $A(u)$, $A(x)$ and $A(\lambda)$ spanned by the bases $\{\psi_1,\ldots,\psi_{m_1}\}$, $\{\xi_1,\ldots,\xi_{m_2}\}$ and $\{\zeta_1,\ldots,\zeta_{m_3}\}$ respectively. A parametrization technique restricts one or more of the variables u, x and λ to their associated finite-dimensional subspaces; the corresponding parametric expressions are

$$u = \sum_{i=1}^{m_1} \alpha_i \psi_i(t) \ , \tag{B.9}$$

$$x = \sum_{i=1}^{m_2} \beta_i \xi_i(t) \ , \tag{B.10}$$

$$\lambda = \sum_{i=1}^{m_3} \gamma_i \zeta_i(t), \tag{B.11}$$

where ψ_i, ξ_i and ζ_i are scalar-valued basis functions and α_i, β_i and γ_i are coefficient vectors of appropriate dimensions.

(1) Control Parametrization

The Lagrangian statement of this procedure is to determine the triple $(\bar{x},\bar{u},\bar{\lambda})$ such that

$$L(\bar{x},\bar{u},\bar{\lambda}) = \sup_{\lambda\varepsilon A(\lambda)} \left[\inf\{L(x,u,\lambda) \mid u\varepsilon S(u), x\varepsilon A(x)\}\right] . \tag{B.12}$$

This can be seen to be equivalent to its usual statement of determining $\bar{u}$ such that

$$J(\bar{u}) = \inf\{J(u) \mid u\varepsilon S(u)\} \tag{B.13}$$

subject to the state equation (B.1). The gradient of J with respect to the parameter α_i can be shown to be given by

$$\frac{\partial J}{\partial \alpha_i} = \int_0^T \frac{\partial H}{\partial u} \psi_i \, dt \ , \qquad i = 1,\ldots,m_1 \ , \tag{B.14}$$

where the co-state λ is given by equation (B.4)

(2) Ritz-Trefftz

The Lagrangian statement of this procedure (see [1]) is to determine the triple $(\bar{x},\bar{u},\bar{\lambda})$ such that

$$L(\bar{x},\bar{u},\bar{\lambda}) = \sup_{\lambda\varepsilon S(\lambda)} \left[\inf\{L(x,u,\lambda) \mid u\varepsilon A(u), x\varepsilon A(x)\}\right] . \qquad (B.15)$$

This is equivalent to maximizing the Lagrangian L with respect to the parameters $\gamma_1,\ldots,\gamma_{m_3}$ subject to the conditions (B.4), (B.5) and the initial condition $x(0) = x_o$. The gradient of L with respect to each parameter γ_i is given by

$$\frac{\partial L}{\partial \gamma_i} = \int_0^T (-\dot{x} + f)\zeta_i \, dt , \qquad i = 1,\ldots,m_3 , \qquad (B.16)$$

where λ is given by (B.11) and x, u are computed from (B.4), (B.5).

We note that when the Ritz-Trefftz algorithm is applied to the linear quadratic problem using patch bases, the Lagrangian takes the form

$$L = \int_0^T \langle z, Gz\rangle \, dt , \qquad (B.17)$$

where G is a positive definite matrix possessing a band structure. This is an attractive computational feature of the algorithm.

(3) Trajectory Approximation

For this procedure the state and co-state are restricted to $S(x)$ and $S(\lambda)$ respectively in its Lagrangian statement. This is equivalent to parametrizing x and λ according to (B.10) and (B.11); the control u is determined from the equation (B.5). The gradients of L with respect to the parameters β_i and γ_i are given by

$$\frac{\partial L}{\partial \beta_i} = \int_0^T (\dot{\lambda} + \frac{\partial H}{\partial x})\xi_i \, dt , \qquad i = 1,\ldots,m_2 , \qquad (B.18)$$

and

$$\frac{\partial L}{\partial \gamma_i} = \int_0^T (-\dot{x} + f)\zeta_i \, dt , \qquad i = 1,\ldots,m_3 , \qquad (B.19)$$

subject to the conditions $x(0) = x_o$ and $\lambda(T) = 0$.

We remark here that the formulation of the trajectory approximation procedure in [2] requires the approximate solution to satisfy the conditions

$$\int_0^T (\dot{\lambda} + \frac{\partial H}{\partial x})\zeta_i \, dt = 0 , \qquad i = 1,\ldots,m_3, \tag{B.20}$$

$$\int_0^T (-\dot{x} + f)\xi_i \, dt = 0 , \qquad i = 1,\ldots,m_2 , \tag{B.21}$$

which are obviously not the same as the optimality conditions obtained by setting the gradients in (B.18) and (B.19) to zero.

(4) Ritz-Galerkin

In this procedure the control, state and co-state are restricted to $S(u)$, $S(x)$ and $S(\lambda)$ respectively in its Lagrangian statement. In other words, u, x and λ are parametrized according to (B.9), (B.10) and (B.11) and the sup-inf operation is performed on the Lagrangian with respect to the parameters α_i, β_i and γ_i. The gradients of L with respect to these parameters are given by

$$\frac{\partial L}{\partial \alpha_i} = \int_0^T \frac{\partial H}{\partial u} \psi_i \, dt , \qquad i = 1,\ldots,m_1 , \tag{B.22}$$

$$\frac{\partial L}{\partial \beta_i} = \int_0^T (\dot{\lambda} + \frac{\partial H}{\partial x})\xi_i \, dt, \qquad i = 1,\ldots,m_2 , \tag{B.23}$$

and

$$\frac{\partial L}{\partial \gamma_i} = \int_0^T (-\dot{x} + f)\zeta_i \, dt , \qquad i = 1,\ldots,m_3 , \tag{B.24}$$

subject to $x(0) = x_o$ and $\lambda(T) = 0$.

REFERENCES

[1] W. E. Bosarge, Jr. and O. G. Johnson: Direct Method Approximation to the State Regulator Problem Using a Ritz-Trefftz Suboptimal Control, IEEE Trans. Automatic Control, AC-15 (1970) pp.627-631.

[2] L. L. Lynn, E. S. Parkin and R. L. Zahradnik: Near-Optimal Control by Trajectory Approximation, Ind. Engng. Chem. Fund., Vol.9 (1970), pp.58-63.

APPENDIX C

SOLUTION OF A BOUNDARY CONTROL PROBLEM

Introduction

In Chapter 6 we described the SP procedure for solving optimal control problems in distributed parameter systems involving control variables that are distributed. However, many problems arising from practical applications involve controls which are allowed to act only at the physical boundary; these are known as boundary control problems. Concrete examples of such problems are found in [1].

As pointed out earlier, the SP method of Chapter 6 is not applicable to these boundary control problems. In these cases the CP procedure is still applicable; it is the purpose of this appendix to illustrate the CP procedure by applying it to a specific boundary control problem.

Problem Formulation

The problem that we consider here involves the following heat conduction system (see [2])

$$\frac{\partial x}{\partial t}(y,t) = \frac{\partial^2 x}{\partial y^2}(y,t), \qquad (y,t) \in [0,1] \times [0,1]. \tag{C.1}$$

The initial and boundary conditions are given by

$$x(y,0) = 0 \tag{C.2}$$

$$\frac{\partial x}{\partial y}(0,t) = 0, \qquad x(1,t) + \frac{\partial x}{\partial y}(1,t) = u(t). \tag{C.3}$$

In this problem x(y,t) is the distributed state variable while u(t) is the boundary control variable acting at the end point y = 1. The objective of the optimal control problem is to find the boundary control u*(t) which minimizes the cost functional

$$J = \frac{1}{2} \int_0^1 \int_0^1 (x-1)^2 \, dy \, dt + 0.05 \int_0^1 (u-1)^2 \, dt. \tag{C.4}$$

Method of Solution

Since the above system is linear, we can determine the Green's function for the boundary control variable u(t) (see [3]). The solution of equations (C.1) - (C.3) is readily found to be

$$x(y,t) = \sum_{k=1}^{\infty} a_k(y) \int_0^t \exp\left[-\beta_k^2(t-\tau)\right] u(\tau) \, d\tau \, , \tag{C.5}$$

where β_k (k=1,2,...) are the roots of the equation

$$\beta \tan(\beta) = 1 \, ,$$

and

$$a_k(y) \equiv \frac{\cos(\beta_k y)}{(\frac{1}{2} + \frac{1}{\beta_k^2}) \cos(\beta_k)} \, .$$

For computational purposes, the infinite series (C.5) must be truncated; that is, only a finite number (p) of terms are retained. Thus we compute x from the expression

$$x(y,t) = \sum_{k=1}^{p} a_k(y) \int_0^t \exp\left[-\beta_k^2(t-\tau)\right] u(\tau) \, d\tau. \tag{C.6}$$

We see that the magnitude of each term in (C.6) depends on the factor $\exp\left[-\beta_k^2(t-\tau)\right]$ which is monotonically decreasing as k increases. As $k \to \infty$, $\exp\left[-\beta_k^2(t-\tau)\right] \to 0$ for $\tau < t$, and since $\beta_{k+1} \simeq \beta_k + 2\pi$, we can

expect reasonable accuracy by retaining only the first few terms.

In solving the problem, we parametrized the control as a parabolic spline over a uniform mesh of the time interval $[0,1]$:

$$u(t) = \sum_{i=1}^{N+2} q_i \psi_i(t) \tag{C.7}$$

where the ψ_i are parabolic B-splines.

From (C.6), we can write

$$x(y,t) = \sum_{i=1}^{N=2} q_i \zeta_i(y,t) \ , \tag{C.8}$$

where

$$\zeta_i(y,t) \equiv \sum_{k=1}^{p} a_k(y) \int_0^t \exp\left[-\beta_k^2(t-\tau)\right] \psi_i(\tau)\, d\tau .$$

Then, using (C.4), the gradient of J with respect to each parameter q is given by

$$\frac{\partial J}{\partial q_i} = \int_0^1 \int_0^1 (x-1)\zeta_i(y,t)dy\, dt + 0.1\int_0^1 (u-1)\psi_i(t)\, dt. \tag{C.9}$$

Numerical Result

Approximate solutions of the problem were obtained by retaining the first 3 spatial modes in the series expansion of (C.6) and using several values of N. The experiment was then repeated for p = 4. The double integrals in (C.6) and (C.9) were evaluated using the 8 × 8 product Gauss-Legendre quadrature formula, while the single integrals were evaluated using the 5th-order Runge-Kutta formula with the integration step size of .05. The minimum cost obtained in each case is included in the following Table.

TABLE C.1 MINIMUM COST

N	$J(\bar{u})$	
	p = 3	p = 4
1	0.1581	0.1581
2	0.1580	0.1580
3	0.1580	0.1580
4	0.1580	0.1580

We see from Table C.1 that the results for p = 3 and p = 4 are practically identical. Moreover, the cost has not decreased by any significant amount as the number of sections N is increased.

In Tables C.2 and C.3 below we summarise the final control and state profiles obtained for the respective cases p = 3 and p = 4 (for both Tables, N = 4). It can be seen that the control profiles for the two cases are practically identical.

Remarks:

The solution procedure as described above depends on the system dynamics being linear. This enables us to obtain the Green's function representation of the state variable in (C.6) with which we then obtain the expression (C.8) for the state as a linear function of the undetermined parameters.

TABLE C.2 STATE AND CONTROL PROFILES (p = 3, N = 4)

t	$\bar{x}(0,t)$	$\bar{x}(.5,t)$	$\bar{x}(1,t)$	$\bar{u}(t)$
0.0	0.000	0.000	0.000	3.24
0.1	0.060	0.129	0.573	2.69
0.2	0.177	0.318	0.716	2.28
0.3	0.324	0.457	0.774	2.00
0.4	0.456	0.564	0.804	1.77
0.5	0.564	0.647	0.822	1.59
0.6	0.650	0.711	0.833	1.44
0.7	0.717	0.760	0.838	1.31
0.8	0.769	0.797	0.838	1.20
0.9	0.808	0.824	0.834	1.10
1.0	0.836	0.841	0.826	1.00

TABLE C.3 STATE AND CONTROL PROFILES (p = 4, N = 4)

t	$\bar{x}(0,t)$	$\bar{x}(.5,t)$	$\bar{x}(1,t)$	$\bar{u}(t)$
0.0	0.000	0.000	0.000	3.24
0.1	0.001	0.126	0.631	2.69
0.2	0.127	0.315	0.765	2.28
0.3	0.280	0.454	0.817	2.00
0.4	0.416	0.561	0.842	1.77
0.5	0.528	0.644	0.856	1.59
0.6	0.618	0.709	0.864	1.44
0.7	0.688	0.759	0.866	1.31
0.8	0.743	0.796	0.864	1.20
0.9	0.784	0.822	0.858	1.10
1.0	0.813	0.840	0.848	1.00

REFERENCES

[1] A. G. Butkovskii: Distributed Control Systems, American Elsevier Inc., N.Y. (1969).

[2] Y. Sakawa: Solution of an Optimal Control Problem in a Distributed Parameter System, IEEE Trans. Automatic Control, AC-9 (1964), pp.420-426.

[3] W. L. Brogan: Optimal Control Theory Applied to Systems Described by Partial Differential Equations, in Advances in Control Systems, ed. C. T. Leondes, Academic Press, N.Y. (1968).